Un langage des sons pour une ingénierie de la perception

Mitsuko Aramaki

▶ **To cite this version:**

Mitsuko Aramaki. Un langage des sons pour une ingénierie de la perception. Acoustique [physics.class-ph]. Aix-Marseille Université (AMU), 2017. tel-01849169

HAL Id: tel-01849169
https://hal.science/tel-01849169

Submitted on 25 Jul 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

AIX - MARSEILLE UNIVERSITE

MEMOIRE D'HABILITATION A DIRIGER DES RECHERCHES

Présenté par

Mitsuko ARAMAKI

Un langage des sons pour une ingénierie de la perception

Soutenance publique le 10 Juillet 2017 à Marseille devant le jury composé de :

Philippe Depalle	Université de McGill, Dpt of Music Research	*Rapporteur*
Jean-Louis Vercher	AMU-CNRS-ISM (UMR 7287)	*Rapporteur*
Isabelle Viaud-Delmon	UPMC-CNRS-STMS (UMR 9912)	*Rapporteure*
Christine Esclapez	AMU-CNRS-PRISM (FRE 2006)	*Examinatrice*
Sylvain Marchand	Université de la Rochelle, L3i (EA 2118)	*Examinateur*
Bruno Torrésani	AMU-CNRS-I2M (UMR 7373)	*Examinateur*

Laboratoire PRISM "Perception, Représentations, Image, Son, Musique" - CNRS - AMU
31, chemin Joseph Aiguier
CS 70071
13402 Marseille Cedex 09

Remerciements

Je tiens tout d'abord à remercier très chaleureusement les membres du jury : M. Philippe Depalle, M. Jean-Louis Vercher et Mme Isabelle Viaud-Delmon, qui ont accepté d'évaluer mon manuscrit, ainsi que Mme Christine Esclapez, M. Sylvain Marchand et M. Bruno Torrésani qui ont accepté de participer au jury.

Les travaux que je décris dans ce document sont le fruit d'un travail collaboratif autour de différents projets impliquant un groupe actif de chercheurs, d'ingénieurs et d'étudiants passionnés sans qui les avancées n'auraient pas été aussi significatives. Je remercie sincèrement les personnes avec lesquelles j'ai eu le plaisir de travailler au quotidien et de partager de beaux moments de sciences. Merci infiniment à Richard Kronland-Martinet et Sølvi Ystad, avec qui j'ai partagé depuis toutes ces années une véritable Aventure scientifique et humaine. J'espère que celle que nous vivons actuellement nous mènera loin. Je souhaite également remercier Charles Verron, Camille Rakovec, Etienne Thoret, Simon Conan, pour la qualité de leur travail de thèse, ainsi que Thomas Bordonné, Antoine Bourachot et Manuel Dias-Alves qui mènent encore leur travail de thèse. Merci à toute l'équipe S2M et en particulier : Thierry Voinier, Olivier Derrien, Charles Gondre, Adrien Merer, Mathieu Barthet, Jocelyn Rozé, Adrien Sirdey, Gaëtan Parseihian, Bertrand Scherrer et Laurent Pruvost.
Je souhaite remercier les collègues de mon ancien laboratoire INCM : Mireille Besson, Jean-Luc Velay, Daniele Schön, Marieke Longcamp, Cyril Magne, Jérémy Danna, Céline Marie, Clément François, et Aline Frey. Un grand merci aux collègues médecins, en particulier Jean Vion-Dury (avec qui je continue l'aventure PRISM), Jean-Arthur Micoulaud-Franchi, Michel Cermolacce et Alexandre El-Kaïm. Merci également à Lionel Bringoux et Christophe Bourdin de l'ISM, Vincent Roussarie et Sébastien Denjean du groupe PSA, Grégory Pallone de Orange Labs, Damián Marelli, ainsi qu'à Loïc Brancheriau et Henri Baillères du CIRAD.

Introduction

Ce document présente mes activités de recherche que j'ai réalisées à partir de la fin de ma thèse jusqu'au début de l'année 2017. Mon parcours scientifique depuis ma formation initiale de physique et d'acoustique m'a conduit à explorer différents axes de recherches centrés sur la relation entre le son et l'humain, en quête d'associer les domaines de l'acoustique et des sciences et neurosciences cognitives. Ils m'ont emmené sur des terrains parfois glissants et imprévisibles mais ils m'ont avant tout conforté dans ma mission de chercheur prônant une vision de la recherche sans frontières, libre et résolument interdisciplinaire. Il me tient à cœur de souligner que les travaux présentés dans ce document ont bénéficié d'un travail collaboratif autour de différents projets impliquant un groupe actif de chercheurs, d'ingénieurs et d'étudiants passionnés sans qui les avancées n'auraient pas été aussi significatives.

Un aspect fondamental de mon cheminement scientifique est également caractérisé par l'interdisciplinarité. Tout au long de mon parcours, j'ai souhaité privilégier l'ouverture vers les disciplines connexes, insatiablement curieuse de nouveaux points de vue. Depuis maintenant plusieurs années, cette orientation interdisciplinaire s'affirme de plus en plus au sein du groupe dans lequel je travaille et le futur proche présage le début d'une aventure scientifique inédite autour de la création d'une unité mixte de recherches dont l'objectif sera de favoriser la transversalité entre sciences, technologies et arts. Affaire à suivre ...

Ce document comporte 6 parties : un CV, mes travaux de recherches, une liste complète des publications, mes activités connexes à mes recherches, le projet de recherche envisagé pour les années futures et une sélection des publications principales. Dans la partie II, j'ai souhaité tout d'abord décrire mon parcours scientifique (chapitre 1), une sorte d'aperçu sur une tranche de dix années de vie scientifique. Mes principaux travaux sont ensuite décrits plus en détail dans les chapitres 2 et 3. Ils sont organisés selon deux aspects qui questionnent la relation entre le son et l'humain : 1) comment le son nous informe ? 2) comment le son influence notre comportement ? Les perspectives de recherche que j'envisage pour les années à venir sont décrites dans la partie V.

Table des matières

Première partie

Curriculum Vitae

Mitsuko ARAMAKI

Affectation **:** Laboratoire Perception, Représentations, Image, Son, Musique (PRISM)
CNRS - Aix-Marseille Univ - FRE 2006
31, chemin Joseph Aiguier
CS 70071
13402 Marseille Cedex 09
Section du Comité National **:** 09
Institut du CNRS **:** Institut des sciences de l'ingénierie et des systèmes (INSIS)

Date et lieu de naissance **:** Née le 3 juillet 1975 à Paris (15ème)
Nationalité **:** Française
Etat civil **:** Célibataire
Courriel **:** aramaki@prism.cnrs.fr

Positions actuelle et précédentes

Janv. 2017 - présent :	Chargée de Recherches (CR1) au Laboratoire Perception, Représentations, Image, Son, Musique (PRISM, FRE 2006, Marseille) dans l'axe "Ingénierie de la perception"
Janv. 2012 - Déc. 2016 :	CR1 au Laboratoire de Mécanique et d'Acoustique (LMA, UPR 7051, Marseille) dans l'équipe "Sons"
Oct. 2010 - Déc. 2011 :	CR1 à l'Institut de Neurosciences Cognitives de la Méditerranée (INCM, UMR 6193, Marseille) dans l'équipe "Langage, Musique et Motricité"
Oct. 2006 - Sept. 2010 :	Chargée de Recherches (CR2) à l'INCM dans l'équipe "Langage, Musique et Motricité"

Formation

Sept. 2004 - Sept. 2006 :	Post-doctorat effectué à l'INCM dans l'équipe "Langage, Musique et Motricité"
Janv. 2004 - Mai 2004 :	CDD effectué au LMA. Contrat à caractère confidentiel.
Juil. 2003 - Déc. 2003 :	Post-doctorat effectué à DIKU, Université de Copenhague (Danemark)
Oct. 1999 - Juin 2003 :	Doctorat de l'université de la Méditerranée (Aix-Marseille II), spécialité Acoustique sur le sujet : *Analyse-synthèse de sons impulsifs : approches physique et perceptive*, co-dirigée par P. Guillemain et R. Kronland-Martinet (LMA)

Production scientifique

- Publications dans des revues avec comité de lecture : 35
- Publications dans des actes de colloque avec comité de lecture : 42
- Ouvrages ou chapitres d'ouvrages : 15
- Edition d'ouvrages : 6

Enseignement, encadrement et animation scientifique

- 2012 - présent : Membre de l'équipe de direction du Master 2 Recherche "Mécanique, Physique et Ingénierie", Spécialité Acoustique, Aix-Marseille Université / Ecole Centrale Marseille.
  Responsable du deuxième semestre et du suivi des stages.
  Depuis 2008 : enseignement dans le module "Analyse, synthèse et transformation des sons" (Resp. S. Ystad)
- 2002 - 2004 : Travaux dirigés et travaux pratiques de Traitement de Signal (Matlab, Electronique) au département "Génie des Télécommunications et Réseaux" de l'IUT d'Aix-en-Provence (site de Luminy).

- Co-direction de thèse de 6 doctorants dont 3 en cours
- Membre du comité d'organisation de 7 conférences internationales
- Membre du comité scientifique de 9 conférences internationales

Valorisation et partenariat industriel

- Responsable scientifique pour le LMA du projet ANR *Physis* (2012-2015)
- Responsable scientifique d'un contrat particulier de recherche externalisée accompagné d'une bourse CIFRE pour la thèse de C. Verron
- Participation aux projets ANR *senSons* (2006-2009), *MetaSon* (2010-2014), *POTION* (2014-2017) et *SoniMove* (2014-2019)

Responsabilités collectives et travaux d'expertise

- Janv. 2017 - présent : Responsable de l'axe "Ingénierie de la perception" du laboratoire PRISM
- Janv. 2012 - Avril 2013 : Membre élue du Conseil de Laboratoire du LMA
- Janv. 2015 : Elue à la responsabilité de l'équipe "Sons" du LMA
- Expertise de projets de recherches pour l'ANR et COST (Suisse)
- Relecture d'articles pour *Computer Music Journal*, *IEEE Transactions on Audio, Speech and Language Processing*, *Journal of the Acoustical Society of America*, *Acta Acustica united with Acustica* et *ACM Transactions on Applied Perception*

Distinctions

- 2013 : Prix du meilleur article du colloque international DAFx'13 pour : "Conan S., Thoret E., Aramaki M., Derrien O., Gondre C., Kronland-Martinet R., and Ystad S. *Navigating in a Space of Synthesized Interaction-Sounds : Rubbing, Scratching and Rolling Sounds*"
- 2012 : Sélection par ACM comme "top best items published in computing in 2012" du livre : "Ystad, S., Aramaki, M., Kronland-Martinet, R. ; Jensen, K. ; and Mohanty, S. (Eds.) *Speech, sound and music processing : Embracing research in India*, LNCS 7172, Springer Verlag Heidelberg"
- 2013 (distinction collective) : Nomination du projet ANR *Physis* comme projet phare.
- 2012 (distinction collective) : Nomination du projet ANR *MetaSon* comme projet phare.

Deuxième partie

Travaux de recherche effectués

Chapitre 1

Parcours scientifique

Le domaine des sciences du son a pris une ampleur remarquable grâce à l'avènement du traitement du signal numérique et des premières synthèses sonores sur ordinateur par M. Mathews dans les années 60 (Mathews, 1963). Près de soixante années de recherche ont permis la construction de modèles et de méthodes aptes à engendrer des sons réalistes ou inouïs et à en étudier leurs corrélats perceptifs. En parallèle, la quête de stratégies de contrôle intuitif pour ces méthodes de synthèse a passionné compositeurs et musiciens depuis ces années. Un grand nombre d'interfaces et de stratégies de contrôle a été proposé, visant pour la plupart des applications musicales. Les avancées technologiques des dernières années ont fortement contribué à la diminution de temps de calcul imposé par l'ordinateur, permettant ainsi l'implémentation d'algorithmes complexes de synthèse en temps-réel. De nouveaux domaines de recherche tels que la réalité virtuelle et le design sonore, ainsi que des nouveaux métiers, se sont également développés grâce à ces avancées technologiques. La synthèse constitue un fort potentiel dans ces domaines et l'aboutissement de ces travaux a ainsi révolutionné le monde académique, mais aussi le monde industriel et économique, en offrant au consommateur des outils technologiques toujours plus innovants et ludiques.

Si un grand nombre de questions restent encore ouvertes à ce jour, notamment sur la compréhension de phénomènes physiques complexes à l'origine de certaines sources sonores, l'attention de la communauté scientifique s'est récemment portée sur des questions relatives à la relation entre le son et l'humain. Parmi ces questions, un des grands défis actuels concerne le problème crucial de la mise en relation entre des morphologies sonores et leurs corrélats perceptifs et cognitifs. En effet, même si les sons produits par synthèse sont d'un réalisme impressionnant, nous ne sommes pas encore capable de totalement comprendre ce qui, dans la morphologie des sons, intervient dans les processus d'attribution du sens. J'ai abordé cette large problématique il y a maintenant une dizaine d'années. Le projet de recherches entrepris à la suite de ma thèse visait ainsi à mieux comprendre la relation entre la physique des sources sonores, la perception des

sons générés et le sens véhiculé par ces sons afin de proposer des outils intuitifs de création de sons dans le contexte de la réalité virtuelle. Ce projet se voulait, et se veut encore maintenant, résolument pluridisciplinaire en élargissant mes perspectives de recherche initiées à la fin de la thèse vers les neurosciences cognitives de l'audition, et plus récemment vers les sciences du mouvement. Cette démarche s'est concrétisée par un post-doctorat obtenu en 2004, puis mon recrutement au CNRS en 2006 sur un poste de chargée de recherches (par la commission interdisciplinaire 45 à l'époque) à l'Institut de Neurosciences Cognitives de la Méditerranée (INCM, CNRS-UMR 6193) à Marseille dans l'équipe "Langage, Musique et Motricité" (resp. M. Besson).

Cette ouverture vers les sciences cognitives m'a notamment permis de reconsidérer la manière de concevoir les outils de synthèse et a mis en évidence la nécessité d'intégrer des concepts perceptifs et cognitifs dès la genèse des processus de synthèse sonore. Avec l'appui sans faille du groupe de recherches pluridisciplinaire qui s'est naturellement constitué au fil des années autour de ces thématiques, un paradigme de contrôle intuitif des sons par synthèse a été développé en accord avec l'approche écologique de la perception proposée par Gibson (1986) dans le cadre de la vision et adaptée à la modalité auditive par McAdams (1993). L'hypothèse sous-jacente à ce paradigme est basée sur l'existence d'invariants morphologiques sonores associés à l'identification perceptive des propriétés des objets (appelés invariants structuraux) et des actions effectuées sur ces objets (invariants transformationnels). Dans ce paradigme, les notions d'objet et d'action peuvent être prises au sens large dans la mesure où, du point de vue du signal, elles sont caractérisées par leur contenu sonore (texture sonore) et de leur évolution temporelle (dynamique) respectivement. Dans ce cadre plus large, elles peuvent alors induire des évocations plus abstraites allant jusqu'à des métaphores inouies (e.g., *gratter le vent*).

Un des enjeux majeurs actuels dans le domaine de la synthèse sonore concerne donc la détermination et le contrôle perceptif de ces invariants morphologiques sonores. Pour ce faire, la méthodologie adoptée a consisté à s'inspirer d'un concept initié par J.C. Risset et D. Wessel et qui a déjà fait ses preuves dans le domaine du son numérique, l'analyse par synthèse (Risset and Wessel, 1999). Elle consiste à déterminer quelle est l'information perceptive importante contenue dans un son en réduisant cette information par la synthèse. Du point de vue de la perception, l'hypothèse d'attribution d'une signification à un son repose en grande partie sur un processus de catégorisation permettant l'association du percept sonore avec une représentation mentale. Les études issus de ma thèse ont contribué au développement d'un synthétiseur de sons d'impact proposant un contrôle haut niveau directement à partir d'une description sémantique des attributs perceptifs de l'objet impacté (matériau, forme et taille perçus, etc.) ainsi qu'un espace de navigation continue dans un espace perceptif de matériaux. Par la suite, avec les travaux de thèse des doctorants que j'ai co-encadré, Charles Verron (thèse CIFRE), Camille Rakovec (thèse MENRT), Etienne Thoret (thèse MENRT), Simon Conan (thèse ANR), et avec le savoir-faire irremplaçable de Charles Gondre (ingénieur d'études en développement informatique), le synthé-

tiseur a évolué vers une véritable plateforme de synthèse des sons de l'environnement que nous avons organisée selon la taxonomie proposée par Gaver (1993) en 3 grandes catégories : les sons d'interactions solidiennes (impact, frottement, roulement, etc.), les sons aérodynamiques (vent, feu, etc.) et les sons de liquide (goutte, vague, pluie, etc). Des considérations physiques et phénoménologiques ont notamment permis de déduire les invariants morphologiques associés à l'évocation de ces sources. La création d'environnements sonores 3D a été abordé spécifiquement dans le travail de thèse de Charles Verron en collaboration avec Grégory Pallone, ingénieur à Orange Labs (Lannion) et Damián Marelli, chercheur à l'Université de Newcastle (Australie). De ce travail, nous avons proposé une stratégie de contrôle de la synthèse permettant de générer des scènes sonores (avec des sources multiples spatialisées) non seulement à partir d'une description sémantique mais également à partir de leur position dans l'espace. L'architecture proposée a permis de répondre efficacement à des contraintes de génération des sons en temps réel imposées dans les applications de réalité virtuelle, car l'exécution est effectuée sans pénalité de calcul. La plateforme a ainsi été progressivement enrichie de modules de synthèse, de paramètres de contrôle intuitif et d'espaces de navigation associés à ces sources sonores. Avec le travail de Charles Verron, Bertrand Scherrer et Laurent Pruvost, qui ont contribué au projet ANR *Physis*, de nouveaux modules de synthèse sonore et les contrôles perceptifs associés répondant à des cas d'usage spécifiques au projet ont été développés dans un contexte audiovisuel et interactif (e.g. éboulements, verre chantant, transitions continues de formes d'objets et de matériaux).

Tous ces outils de synthèse ont par ailleurs instauré des moyens formidables d'interroger la perception humaine. L'expérience acquise durant ces années passées dans un laboratoire de neurosciences m'a permis d'appréhender le phénomène sonore non seulement du point de vue de sa cause physique mais également du point de vue de l'humain, et ce à différents niveaux, i.e. cérébral, cognitif, multimodal, sémiotique. Comme nous le verrons dans les chapitres suivants, la détermination de ces invariants morphologiques sonores permet de construire des modèles de synthèse satisfaisants et capables de restituer des sons réalistes. Toutefois, au-delà de la création sonore, la question fondamentale de savoir comment on reconnaît un son à l'écoute, ou plus généralement comment on attribue une signification à un son perçu reste posée. Cet acte sémiotique repose sur des processus complexes encore largement méconnus, que la physique et l'analyse acoustique ne peuvent naturellement pas expliquer à elles seules. La prise en compte de la dimension perceptive et cognitive pour la synthèse, appuyée par les paradigmes de sciences et neurosciences cognitives, a été une étape déterminante dans la "mise en équation" du phénomène sonore. Cette interdisciplinarité indispensable pour aborder ces problématiques a constitué le cœur du projet que j'ai défendu pour mon concours d'entrée au CNRS, avec une attention

particulière sur cet aller-retour entre les disciplines si difficile à établir. Ainsi, l'association entre les domaines de l'acoustique et des neurosciences a donné lieu à des avancées substantielles dans mon projet de recherche. En maîtrisant la génération des stimuli sonores, il est possible de tester des hypothèses précises sur les processus mis en jeu et d'étendre le concept d'analyse par la synthèse en prenant en compte les données issues de méthodes d'imagerie cérébrale.

Nous nous sommes tout d'abord intéressés aux processus qui sous-tendent la catégorisation perceptive des sons de matériaux impactés produits par le synthétiseur. Les données électrophysiologiques (Potentiels Evoqués, PEs) moyennées en fonction de la classification des participants (Bois, Métal ou Verre) ont mis en évidence des processus cérébraux associés à la perception des sons de Métal qui se différentient de ceux pour les sons de Bois ou de Verre et ce, de manière très précoce, dès 150 ms après le début du son. Cette différentiation observée également sur les PEs successifs supposerait l'implication de processus de traitement de la complexité spectrale (spectres des sons de Métal plus riches que ceux pour le Bois et Verre) et de la durée du son (par conséquent de l'amortissement ; durée plus longue pour les sons de Métal que les sons de Bois et Verre). Ces résultats viennent conforter les invariants morphologiques associés la perception du matériau, en montrant l'importance conjointe de caractéristiques temporels et spectraux du son.

L'expérience suivante a eu pour objectif de mettre en évidence le traitement cérébral du degré de relation sémiotique entre deux sons non linguistiques et en particulier, de déterminer si ce traitement reposait sur la mise en jeu de processus similaires ou différents de ceux mis en jeu dans le langage, en d'autres termes, attribue-t-on un sens aux sons de la même manière que l'on attribue un sens aux mots ? Cette question centrale et toujours débattue dans le domaine des neurosciences a été abordée le plus souvent à propos de la comparaison entre le langage et la musique. De nombreuses études ont consisté à mettre en évidence des modifications anatomo-fonctionnelles d'un "cerveau musicien" afin de révéler l'influence d'une expertise musicale vers d'autres fonctions cognitives. Ces problématiques ont constitué un axe de recherche important au sein de l'équipe dans laquelle je travaillais à l'INCM. En collaboration avec Céline Marie et Aline Frey à l'époque en thèse et en post-doctorat respectivement, nous avons comparé des effets d'amorçage conceptuel entre des scènes sonores d'environnement et/ou des stimuli linguistiques (mots parlés). Les méthodes d'analyse-synthèse développées précédemment ont assuré la génération de stimuli parfaitement calibrés pour ces expériences. Les résultats ont montré la mise en jeu de processus cérébraux communs aux deux types de stimuli, notamment des temps de réaction et des taux d'erreurs ont plus longs/plus grands et au niveau des PEs, une négativité fronto-centrale autour de 450 ms après le début du stimulus cible de plus grande amplitude pour les cibles ambiguës que pour les cibles reliées. Les résultats ont également montré des différences topographiques de l'activité cérébrale qui mettent en évidence des spécificités propres à la nature des stimuli sonores. Acoustique et neurosciences n'ont ainsi pas encore révélé tout

leur potentiel. En effet, cette interdisciplinarité promet des voies d'explorations sonores inédites que je développerai dans mes perspectives de recherche.

Si le contenu informatif des sons est indéniable, leur action sur le comportement humain a été beaucoup moins étudiée dans la littérature, la principale raison étant liée aux difficultés d'un contrôle calibré des stimuli sonores. Durant ces dernières années, mes recherches se sont orientées vers la construction de sons pouvant être utilisés non seulement pour informer mais également pour guider vers une cible (au sens large) et même susciter un comportement moteur particulier chez l'auditeur. Ces recherches s'appuient sur un point de vue théorique qui suppose que notre perception est essentiellement un processus actif : nos actions sont régulées par notre perception, et notre perception dirige nos actions (boucle perception-action). De plus, les processus de perception s'avèrent être fortement dépendants du contexte lié à l'environnement et se basent sur des processus d'intégration globale des informations (auditives, visuelles, sensorimotrices, etc.). Les études que nous avons alors menées ont permis de tester la validité et la robustesse des invariants morphologiques sonores dans un contexte multimodal et de définir des stratégies de contrôle de ces invariants pour une tâche donnée.

Les premières études sur les interactions son/geste humain que nous avons menées ont été effectuées au cours de la thèse d'Etienne Thoret, en collaboration avec Jean-Luc Velay et Jérémy Danna, chercheurs au LNC (Laboratoire de Neurosciences Cognitives, Marseille) ainsi qu'avec Christophe Bourdin et Lionel Bringoux, enseignants-chercheurs à l'ISM (Institut des Sciences du Mouvement, Marseille). Ces interactions son/geste ont été étudiées en incorporant des connaissances propres aux sciences du mouvement au sein des modèles sonores. Grâce à la synthèse et aux expériences menées, les propriétés acoustiques sous-jacentes à la perception d'un son évoquant une dynamique spécifique au geste humain, ont été identifiées aussi bien du point de vue purement sonore que dans un contexte multimodal mettant en jeu l'audition, la vision, et la motricité. Nous avons tout d'abord montré qu'il était possible de percevoir un geste graphique produit par une personne qui dessine, uniquement à travers le son évoquant le frottement entre son stylo et la surface, en particulier à travers les variations de timbre. Les résultats d'une série d'expériences ont en effet révélé que le profil de vitesse pouvait s'identifier à un invariant morphologique du type transformationnel. Ce profil transmet alors, via le son produit, des informations pertinentes sur le geste sous-jacent (évocation d'un geste fluide ou saccadé) ainsi que dans une certaine mesure, sur la forme géométrique dessinée. Ces résultats sont cohérents avec des études princeps qui ont souligné l'importance du profil de vitesse dans la production motrice d'un mouvement humain (Viviani and Terzuolo, 1982) effectuée à partir d'une perception visuelle ou bien kinesthésique (Viviani et al., 1997).

Nous avons également montré que l'évocation de formes géométriques, en l'occurence la catégorie de l'ellipse, étaient contraintes par des propriétés biomécaniques spécifiques de nos gestes. Ce résultat, notamment obtenu par le biais du son, révèle l'importance des contraintes sensori-motrices dans l'émergence de ce qu'on peut appeler des "prototypes géométriques", concepts centraux de nos représentations mentales qui ont été mis en évidence précédemment seulement par le biais visuel ou haptique (Theurel et al., 2012). Nous avons enfin montré que la reproduction d'un mouvement visuel pouvait être altérée, et même modifiée, par l'ajout d'une dynamique sonore non congruente au mouvement visuel dans une tâche de synchronisation sensori-motrice. Ce résultat majeur apporte un éclairage nouveau sur le rôle et l'importance de la modalité auditive dans l'intégration audiovisuelle d'un mouvement continu, ce qui n'avait jusqu'alors que très rarement été mis en évidence aussi clairement. Cette relation intime entre morphologie sonore et reproduction motrice a été évaluée dans l'activité d'écriture au cours du projet *MetaSon*, les scripteurs dysgraphiques pouvant tirer bénéfice de la sonification de leur geste pour améliorer la coordination spatio-temporelle (Danna et al., 2012a). Ceci instaure ainsi les fondements d'un nouveau paradigme d'étude du son intégrant les interactions multimodales que j'envisage de consolider dans les années futures.

Ces travaux ont été menés avant tout dans un contexte fondamental de recherche mais ils ont également bénéficié des avancées obtenues au travers de nombreuses collaborations à visée industrielle, sociétale et clinique. Ces collaborations ont constitué un lieu de rencontres et d'échanges scientifiques incomparables, tissant des liens entre acoustique, mathématiques, sciences du mouvement, sciences et neurosciences cognitives et phénoménologie, autour de cet objet d'étude commun qui est le son pour l'humain. Les applications qui ont motivé ces collaborations ont permis d'une part, de tester *in situ* la validité des résultats obtenus en laboratoire, et d'autre part, d'initier des problématiques nouvelles pour des études fondamentales en spécifiant des cas d'usage spécifiques. J'ai ainsi eu l'occasion de participer à différents projets financés par l'Agence Nationale de la Recherche (ANR) : *senSons* (coordonné par S. Ystad, LMA), *MetaSon* (coordonné par R. Kronland-Martinet, LMA), *Physis* (coordonné par D. Henry, Audio-Gaming) dans lequel j'ai été responsable scientifique pour le LMA, et *SoniMove* (coordonné par R. Kronland-Martinet, LMA), ce projet est toujours en cours. Le projet *senSons* m'a donné l'opportunité d'effectuer un post-doctorat à l'INCM. Les travaux menés ont consolidé les fondations pluridisciplinaires de notre paradigme d'étude basé sur l'existence d'invariants morphologiques sonores porteurs de sens. Par la suite, le projet *MetaSon* a abordé le problème générique de l'utilisation optimale des sons dans le contexte de la réalité augmentée à travers deux applications spécifiques, les sons pour les véhicules silencieux du futur et l'aide à l'apprentissage par le son de l'écriture chez les enfants dysgraphiques. Ce projet a donné lieu aux prémices de la constitution

d'un véritable langage des sons, axe de recherche que je souhaiterais poursuivre dans le futur. Le projet *Physis* a permis la mise en situation des modèles de synthèse dans un contexte audiovisuel et surtout interactif et a montré le potentiel de la synthèse pour les jeux vidéo. Par la suite, le son comme vecteur d'influence sur l'humain constitue la problématique majeure du projet *SoniMove*. Ce projet vise à définir un apport optimal du son pour des applications industrielles et sociétales majeures : le développement d'interfaces homme-machine innovantes basées sur le son pour l'automobile, l'ajustement ou l'apprentissage de gestes experts dans le sport (golf) et la musique, ainsi que l'aide à la régulation posturale. Par ailleurs, dans un contexte de facture instrumentale, nous avons mené une étude sur la relation entre les propriétés acoustiques de 59 essences de bois (tropicaux et sub-tropicaux) et la perception du timbre des sons de ces bois engendrés par percussion, en collaboration avec Henri Baillères et Loïc Brancheriau, chercheurs au Centre de coopération Internationale en Recherche Agronomique pour le Développement (CIRAD, Montpellier). L'objectif consistait à mieux comprendre comment un luthier expertisait ses échantillons de bois pour la fabrication de xylophones et à prédire son classement perceptif afin de proposer à terme des essences de bois alternatives à celles utilisées actuellement (Palissandre, Padouk, etc.) qui produiraient des sons de qualité sonore supérieure. Les résultats de cette étude, en particulier les descripteurs acoustiques mis en évidence, ont contribué à la définition des morphologies sonores pour les sons d'impact.

Enfin, les applications cliniques ont fait l'objet d'une collaboration fructueuse avec les Dr. Jean Vion-Dury, Dr. Michel Cermolacce, Dr. Jean-Arthur Micoulaud-Franchi et Dr. Alexandre El-Kaim, médecins psychiatres du Pôle de Psychiatrie Solaris du CHU de Ste Marguerite (Marseille). L'ensemble des études menées ont été initiées sur la constatation clinique générale que les pathologies psychiatriques (i.e., schizophrénie ou trouble bi-polaire) ou cérébrales (i.e., traumatisme crânien ou encéphalite) sont souvent associées à des anomalies de l'écoute des sons et de la musique. Au cours des travaux de thèse de sciences de Jean-Arthur Micoulaud-Franchi et de master d'Alexandre El-Kaim dont j'ai pu suivre les avancées de près, des outils d'exploration de l'écoute des sons auprès de patients souffrant de schizophrénie ont été développés. Nous avons tout d'abord mis en évidence dans la modalité auditive des modifications de l'organisation perceptuelle qui se sont traduites par des difficultés de catégorisation des sons de l'environnement (sons d'impact) et par une perception modifiée de la familiarité et de la bizarrerie indiquant une tendance d'ambivalence de l'écoute sonore, tendance révélée par ailleurs dans des tâches de reconnaissance d'émotions (Trémeau et al., 2009). Nous avons également validé un dispositif d'écoute en temps réel comme outil de mesure de l'envahissement perceptuel, une des anomalies perceptives majeures reconnues dans la schizophrénie (McGhie and Chapman, 1961). Les résultats obtenus ont permis de confirmer la présence d'un sentiment d'envahissement plus important chez les patients que chez les témoins et que cet envahissement était corrélé à un défaut de filtrage sensoriel mesuré par la méthode des PEs (paradigme P50). Ce dispositif étant fondé

sur l'utilisation du synthétiseur de sons d'environnement, nous avons montré l'intérêt de la synthèse sonore pour l'étude des troubles de la perception, en particulier dans l'utilisation de sons complexes, écologiques et contrôlables. A terme, l'ensemble de ces études vise à contribuer à la réalisation d'un test global de diagnostic clinique mais également de remédiation par le son qui pourra être généralisé aux patients atteints d'autres pathologies psychiatriques. Pour aborder l'aspect remédiation par le son, nous envisageons de repenser les procédés de stimulation auditive par bio- et/ou neurofeedback afin d'agir de manière interactive et personnalisée, et par conséquent plus efficacement, sur les stratégies mentales développées par les patients. Dans un premier temps, nous souhaiterions nous intéresser aux troubles avec déficit attention et hyperactivité (TDAH) et de troubles du sommeil (insomnie chronique) pour lesquels les thérapies classiques de neurofeedback sont déjà mises en place.

Pour compléter cet aperçu, mon parcours scientifique a été jalonné d'étapes régulières comme le cycle de conférences CMMR – Computer Music Multidisciplinary Research. En effet, je suis membre du "steering committee" de ce cycle de conférences dont la première édition a eu lieu à Montpellier en 2003. Dans ce contexte, j'ai eu l'opportunité de participer activement à l'organisation de ces congrès à partir de l'édition 2009, soit en tant que membre du comité scientifique, Paper ou Proceeding chair. Ces conférences donnent lieu depuis le début à la publication annuelle d'un livre de la série LNCS Lecture Notes in Computer Science (édition Springer Verlag Heidelberg) rassemblant des articles scientifiques issus des travaux présentés dans la conférence, et dont j'ai eu en partie la charge éditoriale.

La suite de ce document présente plus en détail les travaux et résultats brièvement décrits dans le décours de mon parcours scientifique. Ils sont organisés selon deux questions fondamentales sur la relation entre le son et l'humain : 1) comment le son nous informe ? 2) comment le son influence notre comportement ? Ces questions centrées sur le son ne sont pas indépendantes. Les études sur le contenu informatif des sons constituent une base nécessaire de connaissances pour aborder les aspects relatifs à l'utilisation du son comme moyen de guidage ou d'influence sur le comportement humain.

Chapitre 2

Informer par le son

Notre environnement sonore nous “parle” constamment. De manière presque inconsciente, nous traitons en permanence les flux d’informations auditives, combinées avec celles des autres modalités. Une thématique importante de ma recherche a consisté à mieux comprendre les caractéristiques acoustiques qui sous-tendent l’attribution d’un sens à ces sons d’environnement. Les études menées ont permis d’une part, la construction de modèles de synthèse permettant de créer des sons dont le contenu informatif est explicite et d’autre part, d’assurer leur contrôle à haut-niveau via une description de la perception et du ressenti évoqué par ces sons. Sur la base d’un paradigme d’étude appelé {action-objet}, un synthétiseur de sons d’environnement a été développé, rendant possible l’hybridation de ces morphologies à l’infini et la création de métaphores sonores évoquant même des situations non réalisables dans la réalité. Les outils ainsi mis en œuvre donnent accès à une infinité de palette de timbres et de dynamiques exploitables pour de nombreuses applications dans le domaine industriel ou médical.

2.1 Détermination d’invariants morphologiques sonores

La question de la reconnaissance auditive de sources de notre environnement a fait l’objet de nombreuses études. Sur la base de la théorie écologique de la perception (Gibson, 1986), les études menées par McAdams (1993) ont suggéré l’existence d’indices sonores contenus dans le signal acoustique et qui sont responsables de l’identification perceptive des sources sonores. Elles ont conduit à la définition de deux types d’invariants appelés *invariants structurels* et *transformationnels* liés respectivement aux propriétés perceptives de l’objet et de son interaction avec l’environnement. En accord avec Gaver (1993), les sons de l’environnement peuvent être classés en fonction de la nature des interactions qui sont mis en jeu, à savoir : les sons issus d’interactions solidiennes (impacts, frottement, roulement, etc.), les sons aérodynamiques (vent, feu, etc.) et les sons produits par des liquides (goutte, vague, pluie, etc.). Cette classification suggère une forte relation entre la cause physique et les évocations qui en sont déduites. Même si elles n’ex-

pliquent pas totalement notre perception, des considérations physiques et phénoménologiques peuvent nous donner des informations et des conjectures importantes pour la détermination de ces invariants morphologiques (e.g., Giordano and McAdams, 2006; Rath and Rocchesso, 2005). Nous verrons que c'est le cas pour les sons d'interactions continues (par exemple, le roulement). Par ailleurs, lorsque ces considérations atteignent leur limite, de nombreuses études ont montré que ces invariants peuvent être extraits de considérations perceptives (McAdams, 1993) ou basées sur l'analyse acoustique et mathématique du signal (e.g., Kronland-Martinet et al., 1997).

La méthodologie pour mettre en évidence ces morphologies sonores a ainsi consisté à confronter ces points de vue physique, phénoménologiques, acoustiques, mathématiques et perceptifs en fonction des catégories des sons de l'environnement. En outre, le concept d'analyse par synthèse permet de réduire la quantité d'informations contenue dans le signal à ces morphologies sonores et par conséquent de tester leur validité perceptive.

2.1.1 Evocation de l'objet

Perception du matériau

Les travaux issus de ma thèse ont conduit à la détermination d'invariants morphologiques sonores associés à l'évocation des matériaux. Des études existantes basées sur des tests psychoacoustiques ont permis de montrer que l'indice le plus important pour la reconnaissance du matériau, pour des sons de type impulsif est l'amortissement (temps de décroissance des composantes spectrales). Ainsi, l'amortissement permet de différencier, par exemple, un son produit par une structure métallique (son résonant) d'un son produit par une structure boisée (son amorti). Je me suis intéressée en premier lieu à la perception du matériau à travers une expérience de catégorisation sonore de différents matériaux frappés (Bois, Métal et Verre). J'ai fait appel aux outils d'analyse-synthèse développés durant ma thèse pour la construction de stimuli rigoureusement calibrés et à une technique de "morphing sonore" pour la construction de continua de sons (simulation des transitions Bois-Métal, Bois-Verre et Verre-Métal). Le concept de continuum permet de déterminer jusqu'à quel point les caractéristiques physiques d'une source sonore peuvent être modifiées sans perturber sa catégorisation. Sur la base des données comportementales, deux types de sons ont été défini : les sons "typiques" de chaque catégorie (définis comme les sons classés par plus de 70% des participants, situés aux positions extrêmes des continua, voir Figure 2.1) et les sons ambigus (classés par moins de 70% des participants, situés aux positions intermédiaires). L'analyse acoustique des sons typiques a consisté à évaluer les principaux descripteurs de signaux connus pour leur importance dans la caractérisation des sons d'impact (amortissement) et plus généralement dans la caractérisation du timbre (i.e., centre de gravité spectral, temps d'attaque, étendue spectral, flux spectral, rugosité). Cette analyse a permis de confirmer l'importance de l'amortissement, largement démontré dans la littérature,

mias aussi de mettre en évidence l'importance de la rugosité. Du point de vue du synthétiseur, la calibration des paramètres d'amortissement a permis de proposer un espace de contrôle intuitif du matériau perçu directement à partir d'un label (Bois, Métal ou Verre ; cf. section 2.2).

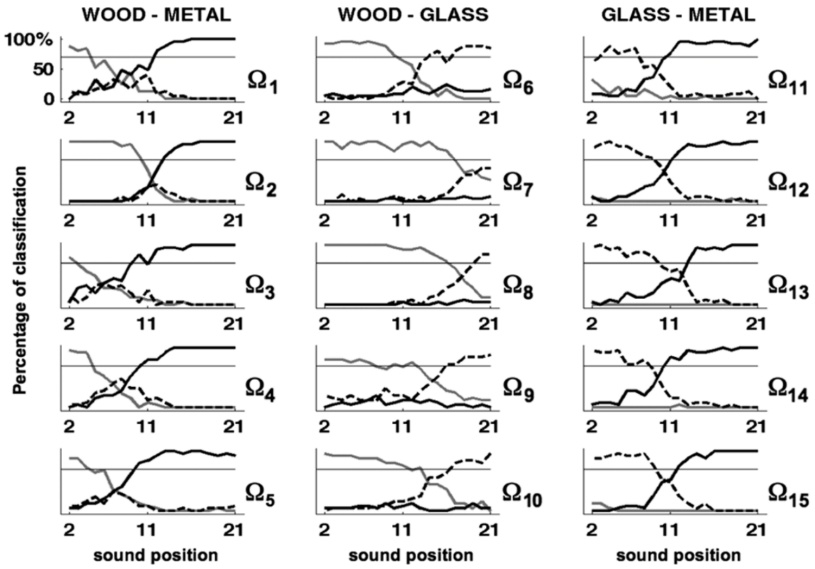


FIGURE 2.1 – Pourcentage de classification des sujets dans la catégorie Bois (gris), Métal (noir) et Verre (pointillés) pour chaque son des 15 continua sonores Ω_i. Le seuil de 70% est représenté par la droite horizontale grise. Figure extraite de (Aramaki et al., 2011).

Classification perceptive : évaluation des invariants

Dans l'idée de réduire la quantité d'information contenue dans un signal acoustique à ses attributs fondamentaux liés à son timbre, une approche statistique a été explorée en complément de ces études perceptives. Dans une première étude, l'objectif consistait à définir un ensemble de classifieurs permettant de catégoriser automatiquement les sons d'impact selon le matériau perçu (Bois, Métal et Verre). Même si le problème de la création des sons est différent de celui de leur classification, la validité des descripteurs acoustiques qui catégorisent le mieux les sons constituent un point de vue pertinent pour la détermination de nos invariants morphologiques ainsi que de leur contrôle. Dans le cadre de la norme standardisée MPEG, une large palette de descripteurs (temporels, spectraux, spectro-temporels) a été définie. Cette norme se base sur le calcul de milliers de descripteurs, toutefois cela reste moins coûteux que de considérer le signal lui-même. Ainsi, au travers des calculs de descripteurs, le signal peut être classifié selon des taxonomies pré-définies telles que les genres musicaux, le type de sons d'environnements, la parole, etc. De manière plus générale, le calcul de descripteurs est très largement utilisé dans les domaines où une masse importante de données ont besoin d'être classifiées, en anglais on parle alors de "data retrieval" ou encore d' "information retrieval". Notre ambition était alors

de proposer une méthode de classification basée sur un nombre réduit de descripteurs et sur des critères perceptifs et non physiques. Le choix des descripteurs est donc déterminant pour la validité des modèles. Nous avons choisi ceux dont on connaît la pertinence d'un point de vue perceptif, en particulier les descripteurs de timbre. Ces modèles statistiques basés sur la régression logistique binaire ont été calibrés et validés avec les sons jugés typiques des catégories Bois, Métal et Verre issus de l'expérience précédente de catégorisation. En accord avec la littérature et les résultats précédents, l'amortissement s'est révélé être le prédicteur le plus important dans les modèles de Bois et de Métal. Le modèle de Verre s'est essentiellement basé sur les descripteurs spectraux comme la rugosité ou l'étendue spectrale. L'ensemble de ces modèles ont affiché un taux de classification correcte supérieur à 80% (Aramaki et al., 2009b).

Cette étude menée en collaboration avec L. Brancheriau et H. Baillères, chercheurs au Centre de coopération Internationale en Recherche Agronomique pour le Développement (CIRAD, Montpellier), nous a par ailleurs permis de conduire une autre étude sur la relation entre les propriétés mécaniques, acoustiques et perceptives de 59 essences de bois (tropicaux et sub-tropicaux) sur la base d'analyse des sons générés par percussion. L'objectif consistait à mieux comprendre comment un luthier expertisait ses échantillons de bois pour la fabrication de xylophones et à prédire son classement perceptif pour la facture instrumentale. Le coefficient d'amortissement du son est apparu comme paramètre pertinent pour prédire l'expertise du luthier et un modèle de classification non linéaire a été établi (corrélation de $R^2 = 0.82$ entre la classification observée et prédite) (Aramaki et al., 2007a, 2006a). Elle ouvre également des perspectives pour affiner le modèle mécanique d'amortissement de sorte qu'il soit valide pour la perception et à terme d'élaborer des essences de bois alternatives à celles existantes (Palissandre, Padouk, etc.) sachant que l'amortissement semble fortement corrélé au coefficient de friction interne du matériau (caractéristique de la structure anatomique du bois).

Perception de la forme

Une étude formelle sur la perception des formes de l'objet a été effectuée dans le cadre de la thèse de Camille Rakovec. De la même manière que pour le matériau, nous avons défini des catégories perceptives propres à ces attributs afin de déterminer les invariants responsables des évocations induites. Une première étude a été menée sur la base d'une banque de sons d'impact que nous avons construit à partir d'un grand nombre d'objets du quotidien présentant des types de matériaux et de formes géométriques variés. Un groupe de sujets a ensuite classé tous ces sons en fonction du type de matériau et de forme perçus. Les résultats de ce test perceptif nous ont permis de privilégier des catégories principales de formes perçues, i.e., formes évoquant des forme unidimensionnelle (corde, poutre), bi-dimensionnelle (membrane, plaque) et tri-dimensionnelle pour laquelle la notion de présence/absence de cavité était importante (Rakovec et al., 2013). Afin de déterminer les invariants morphologiques associés, des descripteurs

acoustiques ont été calculés sur les sons jugés typiques de ces catégories de formes. Pour cela, ces sons ont tout d'abord été modélisés sur la base d'une décomposition en somme de sinusoïdes amorties, particulièrement adaptée pour les sons de type impact. En pratique les sons ont été analysés avec la méthode ESPRIT (Estimation of Signal Parameters via Rotational Invariance Techniques) notamment grâce aux travaux de thèse d'Adrien Sirdey (co-direction de R. Kronland-Martinet et O. Derrien, LMA, thèse soutenue le 9 juillet 2013). La mise en œuvre de ces méthodes dites haute-résolution à des sons naturels est souvent délicate car ils s'écartent inévitablement du modèle de signal théorique et le coût de calcul est important. Dans notre cas, l'adaptation de la méthode a été effectuée en appliquant la méthode ESPRIT dans chaque sous-bande de la transformée de Gabor du signal considéré (Sirdey et al., 2011, 2012). Un ensemble de descripteurs de timbre ont été calculés à partir de ces paramètres de signal. Les analyses acoustiques sont en cours et présagent la détermination de caractéristiques propres à la perception des formes ainsi qu'à la mise en place d'une stratégie de contrôle perceptif. Ce travail fait actuellement l'objet des travaux de thèse d'Antoine Bourachot qui démarre cette année (en co-direction avec R. Kronland-Martinet).

2.1.2 Evocation de l'action

Dans le cadre de la thèse d'Etienne Thoret et de Simon Conan, nous avons développé des modèles de synthèse et des stratégies de contrôle associées pour les sons évoquant différents types d'interactions solidiennes : frottement, roulement, grattement et couinement. Un modèle de synthèse efficace, initialement proposé par Gaver (1993) et amélioré par van den Doel et al. (2002), consiste à synthétiser la source de ce type d'interaction continue par une série de micro-impacts qui simule les impacts successifs entre un plectre et les aspérités de la surface parcourue. Sur la base de ce modèle, nous avons montré que les différences perceptives entre les actions "frotter" et "gratter" reposent sur la différence de densité temporelle de ces impacts, i.e. plus la densité d'impacts est grande (respectivement faible), plus le son est associé à l'action frotter (respectivement gratter) (Conan et al., 2012). Pour la perception du roulement, nous avons observé, à partir de simulations numériques basées sur un modèle physiquement informé, que la structure temporelle de la série d'impacts générée suit un motif temporel et fréquentiel spécifique (Conan et al., 2014a). En particulier, les intervalles de temps entre les micro-impacts et les amplitudes associées sont fortement corrélés et suivent des statistiques particulières. Afin de déterminer ces statistiques, un schéma complet d'analyse-synthèse a été proposé. L'analyse a été conduite sur des simulations numériques basées sur un modélisation physique proposée par Rath and Rocchesso (2005). Le processus de synthèse inclut la génération de la série de micro-impacts avec la statistique adéquate, une modulation d'amplitude (favorisant une meilleure perception de la vitesse et de la taille de la bille, Hermes, 1998; Houben, 2002) puis une modélisation de la forme des impacts qui prend en compte le comportement non linéaire de la force d'interaction

(relation durée-amplitude de l'impact suivant une loi en puissance). Cette dépendance semble également être un élément perceptif important dans l'évocation du roulement. La validation de ces invariants morphologiques a été effectuée par des tests perceptifs et la synthèse a permis d'évaluer l'importance relative de ces contributions.

Ces interactions de frottement, roulement, grattement pouvant toutes être modélisées par des suites d'impacts, nous avons alors défini un modèle de synthèse générique permettant d'effectuer des transitions perceptives continues entre ces 3 interactions (Figure 2.2). Ce modèle repose sur les propriétés statistiques des micro-impacts (densités de probabilité $\mathbb{P}$, corrélations entre les impacts $\mathcal{C}$, valeurs moyennes des séries temporelles et d'amplitude μ), un modèle de la forme de l'impact, une modulation d'amplitude (liée à l'évocation de la vitesse et de la taille de l'objet) et un filtre passe-bas contrôlé par la vitesse du geste (lié à l'évocation de la dynamique). Un aspect fondamental de ce modèle est la prise en compte de la dynamique du geste par la vitesse relative entre l'objet et la surface sur laquelle l'objet interagit. En effet, nous verrons plus loin que l'évocation du geste d'un point de vue perceptif repose principalement sur les caractéristiques de ce profil de vitesse.

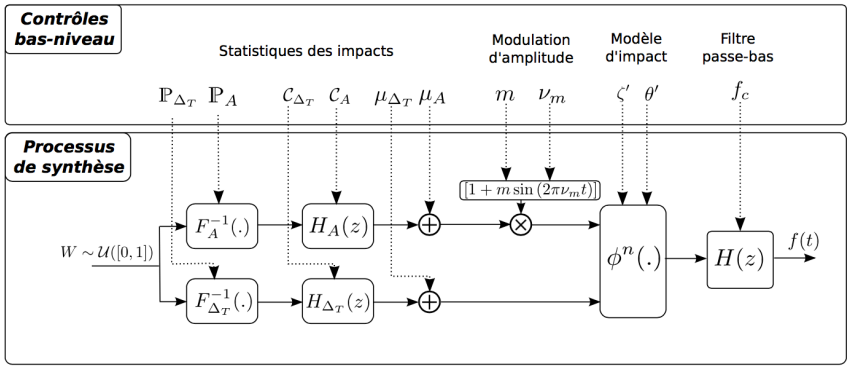


FIGURE 2.2 – Schéma du modèle de synthèse générique des sons d'interactions continues. Figure extraite de (Conan, 2014).

2.1.3 Evocation du mouvement

L'aspect dynamique constitue un attribut fondamental dans la perception des sons. Une manière intuitive de décrire une dynamique est de mimer la trajectoire évoquée et ainsi de se rapprocher du mouvement d'une source sonore dans l'espace. Dans ce contexte, outre certaines considérations physiques, il n'existe que peu de données sur les catégories perceptives relatives à la dynamique perçue d'un son. Dans le cadre du travail de master puis de thèse d'A. Merer (direction de thèse : S. Ystad, LMA, thèse soutenue le 6 juin 2011), nous avons mis en place une série de tests perceptifs via des protocoles de catégorisation (libre et contrainte) et d'évaluation de la dynamique évoquée par un son monophonique (qui permet d'exclure les indices spatiaux

et de se focaliser sur les attributs de timbre) par le biais de dessins. Le protocole d'évaluation par le biais de dessins a consisté à demander aux sujets de dessiner la trajectoire dynamique qui correspondait le mieux à chaque son avec une interface graphique interactive spécialement conçue pour cette étude. Cette interface permet de représenter des trajectoires dynamiques à partir de 9 paramètres accessibles à l'utilisateur (6 paramètres caractérisant la forme de la trajectoire et 3 paramètres caractérisant des variations dynamiques de la trajectoire) et par conséquent, permet de s'affranchir de la variabilité inhérente à des dessins manuscrits. Afin de se focaliser sur les morphologies intrinsèques du son, nous avons adopté une approche basée sur une écoute dite acousmatique initié par Schaeffer (1966) qui permet d'appréhender le son comme une entité propre sans référence à son origine physique (Merer et al., 2010). Ce type d'écoute permet ainsi de s'abstraire de nos références cognitives trop bien repérées (par exemple, un son de voiture ne peut évoquer qu'un déplacement horizontal). En pratique, nous avons travaillé essentiellement avec des sons dits "abstraits", issus de banques de sons de musique électroacoustique, qui ont l'avantage de ne pas évoquer de sources facilement identifiables mais qui peuvent néanmoins véhiculer du sens. Les résultats d'expériences ont permis de définir des grandes catégories perceptives liées à la dynamique (*tomber*, *monter*, *passer*, *s'éloigner*, *tourner*, etc.) et un ensemble de sons les plus représentatifs de chacune de ces catégories. L'analyse des dessins paramétrisés a permis de mettre en évidence des dimensions pertinentes reliées à la trajectoire perçue (*linéaire*, *oscillant* et *circulaire*), à l'orientation globale (*vertical ascendant*, *vertical descendant* et *horizontal*) et à la vitesse (Merer et al., 2007, 2008, 2013). Il est intéressant de noter que la latéralité (mouvement provenant de la gauche ou de la droite) n'a pas été clairement distinguée, ce qui peut s'expliquer par le fait qu'il est difficile de simuler des trajectoires latéralisées avec des sons monophoniques. Ces résultats ont été ensuite validés par synthèse en se basant sur des considérations physiques relatives aux sources mobiles (Chowning, 1971). Ainsi des transformations sonores simulant l'effet Doppler (fréquence), la distance source-auditeur (intensité), l'absorption de l'air (filtrage passe-bas) et la réverbération ont été définies et calibrées pour chaque catégorie de mouvement. Ces trajectoires ont été simulées visuellement avec l'interface utilisée pour le test perceptif précédent. Les sons associés aux trajectoires ont été générés par synthèse. Puis nous avons présenté à des sujets ces trajectoires combinées avec des sons cohérents et incohérents (associés à une autre trajectoire) et nous leur avons demandé de juger de la cohérence perceptive de la scène audio-visuelle. Les résultats présentés sur la Figure 2.3 ont permis de valider des morphologies sonores clairement identifiées à une catégorie de mouvement.

2.2 Contrôle intuitif des sons

La quête de stratégies de contrôle intuitif pour la synthèse a passionné compositeurs et musiciens depuis la réalisation des premiers sons sur ordinateur par M. Mathews (Mathews, 1963).

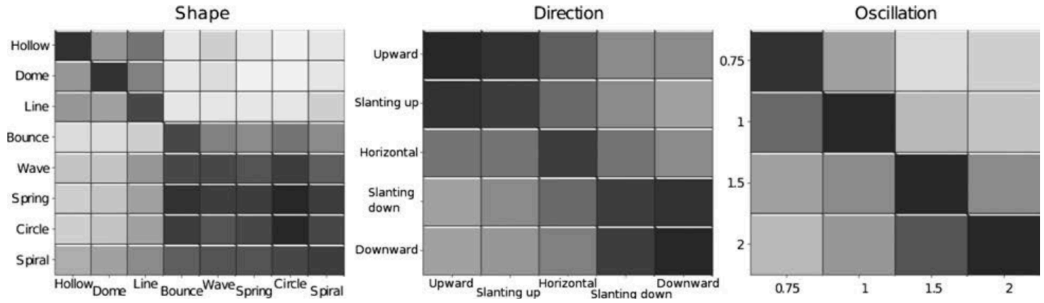


FIGURE 2.3 – Matrices de cohérence entre les trajectoires visuelles (en ligne) et les sons (en colonne). Le degré de cohérence est représenté par le niveau de couleur : la cohérence augmente avec le niveau de noir. Figure extraite de (Merer et al., 2013).

Un grand nombre d'interfaces et de stratégies de contrôle a été proposé, visant pour la plupart des applications musicales (Wanderley and Battier, 2000; Gobin et al., 2003; Moog, 1987). La modélisation physique permet un contrôle basé sur les caractéristiques physiques de la source, cependant elle demande une certaine expertise dans la connaissance des paramètres physiques pour atteindre la qualité sonore souhaitée. Dans le cas de la synthèse par modèle de signal, le contrôle nécessite la manipulation d'un grand nombre de paramètres qui ne sont pas intuitifs pour des non experts. Les travaux de D.L. Wessel constituent une des premières propositions de navigation dans un espace sonore perceptif par synthèse (Wessel, 1976, 1979). Pour cela, il s'est basé sur l'espace de timbre défini par Grey (1977) et a proposé un schéma de contrôle temps-réel basé sur la synthèse additive. Les paramètres de contrôle étaient reliés aux deux premières dimensions de l'espace de timbre : l'enveloppe du spectre d'énergie et le temps d'attaque (ou bien sur la synchronicité temporelle des différentes composantes). Plus récemment, des stratégies de contrôle à partir d'une description verbale du timbre (Gounaropoulos and Johnson, 2006), par machine learning(Jehan and Schoner, 2001), par variations de descripteurs acoustiques (feature-based synthesis (Caetano and Rodet, 2011; Hoffman and Cook, 2006, 2007)) ont été proposées.

Les avancées récentes de l'équipe dans le domaine du contrôle intuitif de la synthèse sonore ont permis de mieux comprendre les relations entre les paramètres de synthèse et les attributs perceptifs relatifs à l'évocation d'une source sonore. En accord avec les invariants présentées dans le paragraphe précédent (section 2.1), nous avons défini un paradigme d'étude appelé {action-objet} dans lequel le son est décrit comme le résultat perceptif d'une action sur un objet. Afin de permettre une utilisation intuitive des modèles de synthèse, nous proposons alors un contrôle par des descriptions sémantiques qui ont pour but de traduire le ressenti de l'auditeur sous forme de mots désignant des attributs perceptifs de la source sonore liés à l'objet et à l'action effectuée. La stratégie de contrôle est fondée sur une architecture à trois niveaux : (a) le haut-niveau utilisant ces descriptions sémantiques associés au ressenti, (b) un niveau in-

termédiaire associé aux caractéristiques acoustiques du signal et par conséquent aux invariants morphologiques sonores (structuraux et transformationnels), (c) le bas-niveau dédié aux paramètres du modèle de synthèse (Figure 2.4). La mise en correspondance entre ces différents niveaux est le mapping. Afin de proposer un contrôle intuitif, la construction de ce mapping a nécessité de prendre en compte la perception des sons dans son aspect le plus large en incluant le point de vue cognitif.

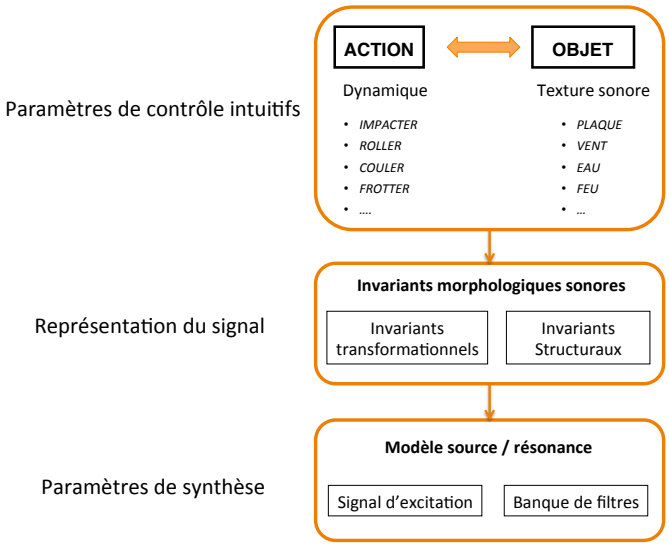


FIGURE 2.4 – Paradigme {action/objet} permettant de générer et contrôler intuitivement des sons et de créer des métaphores sonores en combinant virtuellement des actions et des objets.

Le niveau supérieur donne à l'utilisateur la possibilité de définir les caractéristiques perceptives de l'objet (telles que la nature du matériau, la taille et la forme, etc.) et l'excitation (force, la dureté du marteau, la position d'impact, etc.) à l'aide de descriptions verbales. La Figure 2.5 présente l'interface haut niveau du synthétiseur actuel proposant à l'utilisateur les catégories de sons d'interactions solidiennes (en haut) et de liquides (en bas). La couche intermédiaire est basée sur des descripteurs acoustiques (amortissement, inharmonicité, rugosité, etc.) liés aux invariants morphologiques sonores (Aramaki et al., 2007b, 2008, 2009b). La couche inférieure est constituée de paramètres de synthèse (e.g. coefficient d'amortissement, amplitude et fréquence des composantes).

Ce synthétiseur propose une navigation intuitive dans des espaces sonores d'actions (impact, roulement, frottement, grattement), de matériaux perçus (bois, métal, verre, pierre, plastique) et de formes d'objets (corde, barre, membrane, plaque, coque). La navigation continue à travers ces espaces est basée sur une manipulation des morphologies sonores. Des sons prototypiques sont définis pour chaque catégorie par un ensemble de valeurs particulières des paramètres du modèle, permettant ainsi des transitions continues entre ces prototypes par interpolation de ces paramètres (Aramaki et al., 2006b; Aramaki and Kronland-Martinet, 2006; Aramaki et al.,

2009c,a, 2011; Conan et al., 2013c,b, 2014b).

Pour le roulement, des contrôles haut niveau additionnels ont été proposés dans un module spécifique, tels que les caractéristiques de la bille qui roule (taille, l'asymétrie de forme et vitesse) et l'irrégularité de la surface sur laquelle la bille roule (Conan et al., 2013a). En outre, une méthodologie de contrôle intuitif pour un modèle de son de frottement non linéaire (grincement, couinement etc.) a été proposée. Le mapping entre les descripteurs dynamiques (vitesse, pression) et des paramètres de synthèse est inspiré des diagrammes de Schelleng et permet des transitions cohérentes entre les différents régimes de friction non linéaires.

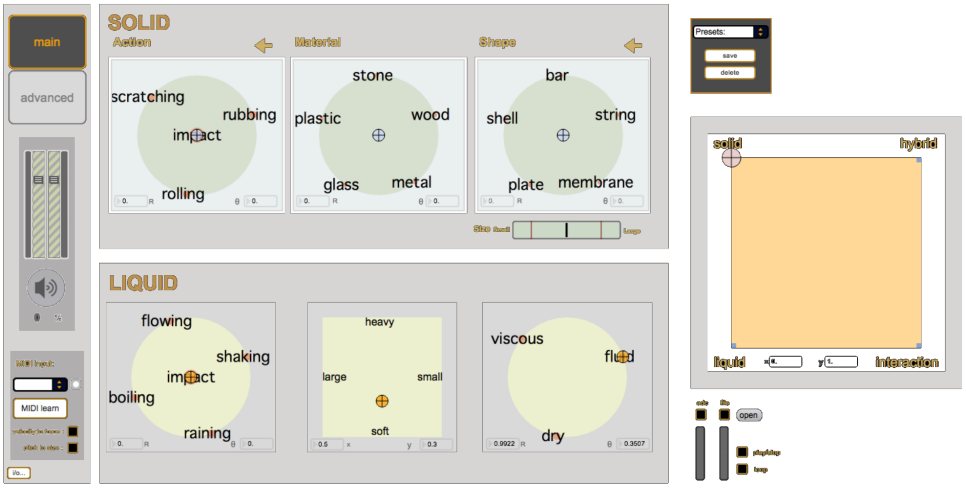


FIGURE 2.5 – Interface haut-niveau du synthétiseur de sons de l'environnement (sons d'interactions solidiennes et sons de liquide) proposée à l'utilisateur.

Métaphores sonores Les notions d'action et d'objet peuvent être prises au sens large, l'action étant associée à une dynamique sonore (évolution temporelle) et l'objet à une texture sonore (contenu timbral). Cette approche est assimilable au concept d'*objet sonore* proposé par Schaeffer (1966) et qui est défini selon deux critères fondamentaux : matière (contenu sonore) et facture (évolution temporelle). En pratique, le paradigme {action-objet} s'adapte naturellement à une approche de la synthèse par modèle source/filtre, où l'information perceptive sur l'objet est contenue dans la partie filtre, et l'information sur l'action dans la partie source. Ainsi en se basant sur le principe de la synthèse croisée, il est possible de faire vivre des textures sonores inertes à la base en appliquant des dynamiques sonores. Les travaux de fin de thèse de Simon Conan abordent spécifiquement la création de métaphores sonores à partir d'hybridations entre des textures stationnaires et les invariants associés à l'évocation de différentes actions. Sur la base d'une série de tests perceptifs, nous avons montré l'avantage de notre méthode par rapport à une approche de design sonore plus classique consistant à additionner les flux sonores (Conan, 2014). La partie droite de la Figure 2.5 (espace de navigation carré) montre le contrôle permet-

tant la création de ces métaphores (en positionnant le curseur sur “hybrid”). La validation de ces métaphores est en cours.

2.3 Création de scènes sonores immersives

Certains sons d’environnement peuvent être larges et diffus en situation naturelle comme le vent, le feu, la pluie ou la vague. Afin de générer de telles scènes sonores, un synthétiseur de sons d’environnement spatialisés a été développé dans le cadre de la thèse de Charles Verron. L’architecture ainsi mise en place donne la possibilité à l’utilisateur d’agir conjointement sur les paramètres acoustiques liés aux sources sonores et sur ceux liés à leurs attributs spatiaux (positionnement dans l’espace 3D). Le contrôle global du synthétiseur est alors effectué de manière intuitive par le biais d’une interface graphique où les sources sonores (choisies parmi un ensemble de sources disponibles : le feu, le vent, la pluie, les vagues, carillons, bruits de pas, interactions solidiennes, etc.) sont placées autour de l’auditeur en définissant la distance et l’étendue spatiale de chaque source (Verron et al., 2010a,b, 2009a,b, 2008a,b, 2007). Par exemple, une scène sonore de pluie peut être créée par une superposition d’éléments constitués de gouttes qui tombent sur le sol de manière aléatoire, de ruissellements plus ou moins localisés et d’un “bruit de fond” réparti tout autour de l’auditeur.

Un critère important a porté sur l’efficacité numérique des modèles de synthèse et des stratégies de contrôle associées, de façon à répondre à la contrainte temps-réel liée à une utilisation interactive du synthétiseur. Contrairement aux procédés actuels de création d’environnements sonores 3D qui comporte deux étapes distinctes (synthèse du son monophonique puis spatialisation), le synthétiseur spatialisé de sons de l’environnement que nous avons développé intègre ainsi une architecture originale combinant un modèle de synthèse additive basé sur une implémentation fréquentielle par transformée de Fourier inverse (FFT-1) et des modules compatibles de spatialisation audio 3D (binaural, transaural ou VBAP, etc.) comme présentée sur la Figure 2.6. Un exemple de stratégie de contrôle global d’une scène complexe composée d’un feu, de vagues et de vent positionnés à différents endroits autour de l’auditeur est présenté sur la Figure 2.7. Les paramètres de contrôle relatifs à la nature de la source sonore et à sa localisation spatiale sont définis au même niveau de la génération du son.

Parmi les sources sonores, les sons bruités ont fait l’objet d’une attention particulière dans ce contexte. En effet, ils correspondent à une importante catégorie de sons d’environnement et la synthèse de ce type de sons présentent des problématiques non encore résolues. Les méthodes de synthèse dans le domaine fréquentiel (comme la méthode par FFT-1) permettent de réduire considérablement le coût de calcul mais présentent des limitations au niveau des résolutions temporelle et fréquentielle. En particulier, la résolution fréquentielle est obligatoirement fixée par la longueur temporelle de la fenêtre d’analyse et de synthèse. Cette contrainte limite l’utilisa-

tion de la méthode par FFT-1 pour la synthèse de sons d'environnement pour lesquels coexistent des signaux transitoires et bruités à bande étroite en fréquence. Pour pallier à cette contrainte, nous avons proposé une généralisation du processus de synthèse basé sur l'algorithme de FFT-1 à un processus de synthèse temps-fréquence basé sur la méthode en sous-bandes (Marelli et al., 2010a,b, 2012; Verron et al., 2010c). Cette technique alternative permet de générer un bruit coloré avec une résolution fréquentielle plus fine que celle inhérente à la synthèse par FFT-1 en prenant en compte les corrélations dans le plan temps-fréquence. Ainsi, contrairement à la synthèse par FFT-1, la méthode en sous-bandes permet la synthèse de sons bruités à bande étroite en fréquence tout en utilisant des fenêtres d'analyse étroites en temps, indispensable pour la synthèse de signaux transitoires.

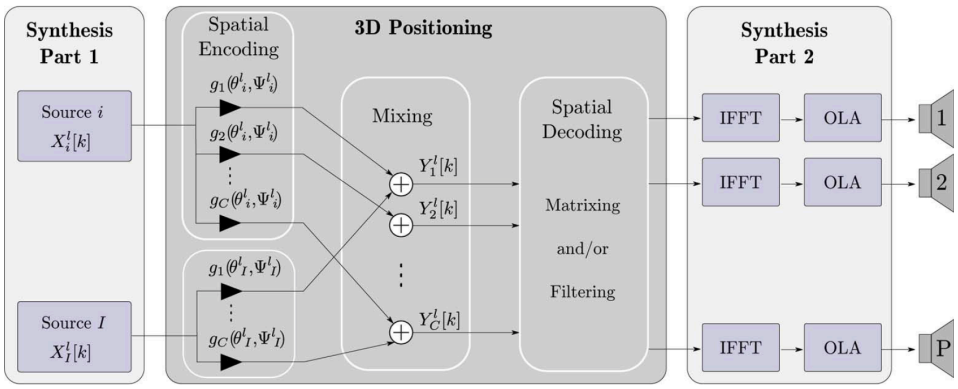


FIGURE 2.6 – Processus de synthèse spatialisée basé sur une implémentation fréquentielle par transformée de Fourier inverse (FFT-1) et des modules compatibles de spatialisation audio 3D. Figure extraite de (Verron et al., 2010a).

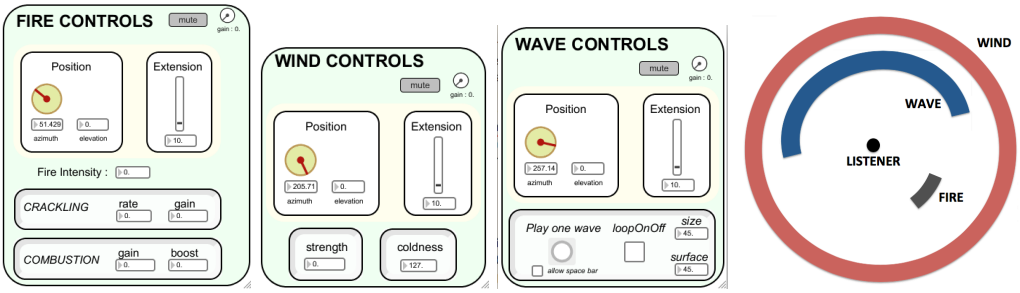


FIGURE 2.7 – Stratégie de contrôle d'une scène ventée sur une plage près d'un feu de camp. Contrôle du positionnement de chacune de ces sources sonores autour de l'auditeur, figure extraite de (Aramaki et al., 2012).

2.4 Bases neuronales d'une sémiotique des sons

La quête d'invariants morphologiques m'a naturellement conduite à m'intéresser aux paradigmes et aux méthodes pratiquées dans le domaine des neurosciences cognitives afin d'approfondir le vaste champ de la sémiotique des sons et de mieux comprendre les processus cérébraux qui sous-tendent le traitement des sons. Les outils de synthèse développés précédemment offrent ainsi des possibilités de façonner finement les stimuli sonores et d'interroger la perception. D'autre part, les corrélats cérébraux et les paradigmes sous-jacents permettront d'apporter des éléments d'éclairage sur le type de sémiose sous-tendu par ce langage des sons. Ainsi, nous avons mis au point différents protocoles expérimentaux afin d'étudier plus en détail le processus de catégorisation perceptive et de traitement sémiotique entre des sons non linguistiques avec la méthode des Potentiels Evoqués (PEs).

Méthode des Potentiels Evoqués (PEs) Cette méthode d'imagerie cérébrale consiste à enregistrer l'activité électrique du cerveau (électroencéphalogramme, EEG) associée au traitement de l'information apportée par un stimulus (sonore, visuel, etc.). Les variations de cette activité électrique synchronisées avec la présentation de ce stimulus sont reflétées par une succession de déflexions positives (P) ou négatives (N) par rapport à un niveau de base appelées composantes qui sont caractérisées non seulement par leur polarité, mais aussi par la latence de leur maximum d'amplitude (par rapport au début du stimulus), par leur distribution topographique sur les différentes positions standardisées des électrodes sur le scalp et par leur signification fonctionnelle. Typiquement, il est communément admis que les premières composantes (P100, N100 et P200) reflètent les étapes de traitement sensoriel et perceptif de l'information et sont obligatoirement évoquées à la suite d'une stimulation. Par contre, la nature des composantes apparaissant après 200 ms dépend des caractéristiques de la stimulation, du protocole expérimental et de la tâche demandée. L'interprétation de ces composantes plus tardives doit alors prendre en compte l'ensemble de ces facteurs.

2.4.1 Processus de catégorisation perceptive

Nous nous sommes ici intéressés aux processus qui sous-tendent la catégorisation perceptive des sons de matériaux impactés issus du synthétiseur. Dans le test de catégorisation présenté en section 2.1, nous avons également mesuré l'EEG des sujets lorsqu'ils effectuaient la tâche. Les données électrophysiologiques moyennées en fonction de la classification des participants (Bois, Métal ou Verre) ont mis en évidence des processus cérébraux pour les sons de Métal différents de ceux pour les sons de Bois ou de Verre et ce, de manière très précoce, dès 150 ms après le début du son (Aramaki and Besson, 2007; Aramaki et al., 2009a, 2011). En particulier, les sons de Métal sont associés à une composante P200 de plus faible amplitude et des composantes N280 et

NSW (Negative Slow Wave) de plus grande amplitude que les sons de Bois et Verre (Figure 2.8, gauche). Sur la base de travaux relatifs aux corrélats acoustiques de ces composantes, on suppose que cette différentiation refléterait des processus de traitement de la complexité spectrale (spectres des sons de Métal plus riches que ceux pour le Bois et Verre) et de la durée du son (par conséquent de l'amortissement ; durée plus longue pour les sons de Métal que les sons de Bois et Verre). Par contre, les résultats n'ont pas révélé de différences significatives sur les composantes P100 et N100 dont on sait qu'elles sont notamment influencées par des variations acoustiques liées au transitoire des sons. Ces résultats semblent appuyer le fait que le processus de catégorisation intégrerait les aspects non seulement temporels mais aussi spectraux du son et par conséquent, viennent conforter les résultats acoustiques précédents montrant l'importance conjointe des descripteurs temporels et spectraux dans la perception du matériau.

2.4.2 Relation sémiotique entre deux sons

L'expérience précédente a permis de mieux comprendre les processus de catégorisation associés au traitement de sons isolés. L'expérience que nous avons mis au point par la suite a eu pour objectif de mettre en évidence un traitement du degré de relation sémiotique entre deux sons et, en particulier, de déterminer si ce traitement repose sur la mise en jeu de processus similaires ou différents de ceux mis en jeu dans le langage (traitement sémantique). Cette question présente un fort intérêt pour instaurer les fondements du langage des sons qui nous souhaitons élaborer. Le traitement des mots du langage est spécifique dans la mesure où les séquences de phonèmes qui forment les mots n'ont pas de signification en elles-mêmes mais acquièrent une signification par le processus de double articulation (de Saussure, 1916). En revanche, il existe une relation intrinsèquement causale entre les sons de l'environnement et la signification que nous leur associons (Ballas, 1993). Ces considérations tendent à conclure que les mots et les sons de l'environnement sont traités différemment par le cerveau. Cependant, des études dans le domaine des PEs ont avancé des arguments en faveur de similarités entre ces 2 processus (Orgs et al., 2006; Plante et al., 2000; Van Petten and Rheinfelder, 1995).

Dans le cadre du travail de master de Céline Marie, nous avons mis en place un protocole d'amorçage (largement utilisé pour le langage) avec des sons linguistiques et non linguistiques sur un même groupe de participants (Aramaki et al., 2010). Pour les sons non linguistiques, les mêmes stimuli que l'expérience précédente de catégorisation sonore de matériaux frappés (section 2.1) ont été utilisés mais présentés cette fois-ci par paires. Les sons linguistiques ont consisté en des mots, pseudo-mots et non-mots et ont été enregistrés spécifiquement pour cette expérience. Pour les deux types de stimuli, les sons de chaque paire différaient par leur degré de relation sémiotique/sémantique : paires reliées (deux sons typiques du même matériau pour les sons non linguistiques ; deux mots pour les sons linguistiques), ambiguës (un son typique suivi d'un son ambigu ; mot suivi d'un pseudo-mot) et non reliées (deux sons typiques de ma-

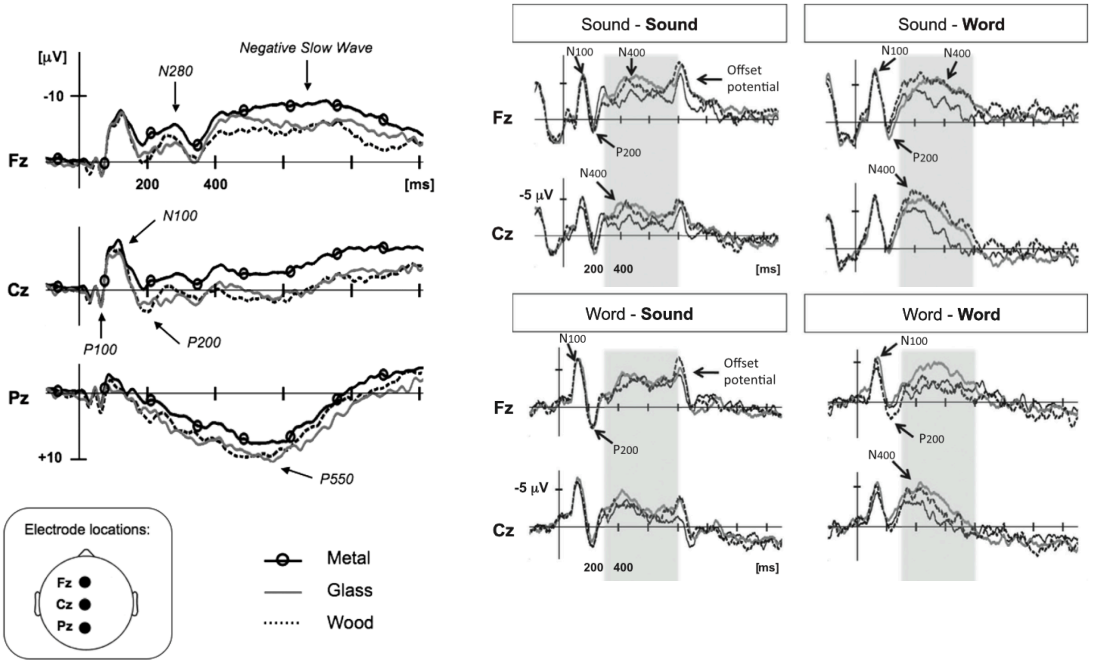


FIGURE 2.8 – **Gauche :** Potentiels Evoqués associés au processus de catégorisation des sons perçus comme du Métal, du Bois et du Verre. Figure extraite de (Aramaki et al., 2011). **Droite :** Effet d'amorçage conceptuel révélé par les Potentiels Evoqués associés au stimulus cible relié (en noir), ambigue (en gris) et non reliée (en pointillé), sur les électrodes fronto-centrales. Figure extraite de (Frey et al., 2014).

tériaux différents ; mot suivi d'un non-mot). Nous avons particulièrement exploré l'effet N400, i.e. la différence d'amplitude sur la composante N400 entre les cibles reliées et non reliées ou ambigües. En effet dans le cas du langage, il a été largement montré que l'amplitude de la composante N400 est inversement proportionnelle au degré du lien sémantique entre un mot et le contexte qui le précède (Kutas and Hillyard, 1984). Les données comportementales ont révélé un pourcentage d'erreurs plus élevé pour les cibles ambiguës que pour les cibles reliées et non reliées pour les deux types de stimuli. Les données électrophysiologiques ont montré des similitudes dans les effets obtenus entre sons linguistiques et non linguistiques, se traduisant par une négativité fronto-centrale autour de 450 ms après le début de la cible plus ample pour les cibles ambiguës que pour les cibles reliées, et une positivité (composante P300) plus ample dans les régions pariétales pour les cibles non reliées. Malgré une différence dans la distribution topographique de ces effets, l'ensemble de ces résultats refléterait plutôt la mise en jeu de processus cérébraux d'amorçage conceptuel communs aux deux types de stimuli.

Une série d'expériences a par la suite été conduite afin d'étudier et de comparer les effets d'amorçage conceptuel entre des scènes sonores d'environnement et des stimuli linguistiques (Frey et al., 2014), en collaboration avec Aline Frey. Les stimuli linguistiques étaient constitués de mots parlés qui ont été associés le plus fréquemment aux scènes d'environnement par des sujets au cours d'un pré-test. Quatre conditions expérimentales entre l'amorce et la cible ont été mis en place : son-son, mot-son, son-mot et mot-mot. Pour chaque condition, l'amorce et la cible différaient par le degré de relation sémiotique : paire reliée, non reliée, et ambiguë, comme pour l'expérience précédente. Dans une première expérience, les sujets devaient évaluer si l'amorce et la cible étaient reliées ou pas (tâche explicite). Dans une deuxième expérience, les sujets devaient se focaliser sur le deuxième stimulus et juger si le stimulus était soit un mot ou pas pour les cibles mot, soit typique d'une scène d'environnement ou pas pour les cibles son (tâche implicite). Dans les 2 expériences et dans les 4 conditions expérimentales, les temps de réaction et les taux d'erreurs ont été plus longs/plus grands et la composante N400 était de plus grande amplitude pour les cibles ambiguës que pour les cibles reliées (Figure 2.8, droite). Ces résultats révèlent l'existence de processus communs de traitement cérébral des scènes auditives et des stimuli linguistiques dans les tâches explicite et implicite. En revanche, des analyses supplémentaires ont révélé des différences entre les expériences et les conditions expérimentales notamment dans la distribution topographique ainsi que dans la durée des effets d'amorçage, reflétant probablement des différences dans les processus d'intégration des attributs perceptifs et cognitifs des sons linguistiques et non-linguistiques.

L'ensemble de ces études basées sur les méthodes d'imagerie conforte l'existence d'un langage des sons. Il reste à mieux cerner les contours de ce langage aussi bien du point de vue théorique que du point de vue de la synthèse.

2.5 Anomalies perceptives

Une collaboration fructueuse avec les médecins de l'Unité Fonctionnelle de Neurophysiologie et Psychophysiologie du Pôle de Psychiatrie Universitaire du CHU de Ste Marguerite, Marseille (les docteurs J. Vion-Dury, M. Cermolacce, J.A. Micoulaud-Franchi et A. El-Kaim) m'a permis d'aborder les aspects pathologiques de l'écoute des sons depuis maintenant plusieurs années. Ces études sont basées sur la constatation clinique que les pathologies cérébrales (i.e., traumatisme crânien ou encéphalite) ou psychiatriques (i.e., schizophrénie ou trouble bi-polaire) sont souvent associées à des anomalies de l'écoute des sons et de la musique. De plus, ces patients présentent souvent des hallucinations acoustico-verbales (voix, bruits, etc.) en période de crise.

La schizophrénie est définie dans le "Diagnostic and Statistical Manual - Revision 4" (DSM IV) par la présence de deux syndromes : délirant (ou productif ou positif) dont font partie les hallucinations, et dissociatif (ou négatif) où figurent désorganisation, repli autistique, rupture

du cours de la pensée. Dans le cadre des travaux de master puis de thèse de sciences de Jean-Arthur Micoulaud-Franchi (direction : J. Vion-Dury, thèse soutenue le 12 Décembre 2013), un premier protocole expérimental a été mis en place et consisté en une version simplifiée du test de catégorisation sonore du matériau perçu (section 2.1) avec un nombre réduit de stimuli. Les résultats ont montré que les patients schizophrènes perçoivent les transitions entre catégories perceptives de manière moins catégorielle que le groupe de sujets témoins, en partie due à une catégorisation différentiée des sons de Métal (Micoulaud-Franchi et al., 2011). Cette différence a été discutée en termes d'anomalies dans la perception du timbre chez les patients schizophrènes rapportées par ailleurs dans la littérature.

Les modifications de l'organisation perceptuelle dans la schizophrénie ont été souvent mises en évidence à travers des retranscriptions d'entretiens. En particulier, la difficulté d'organiser les informations sensorielles, notamment auditives, provenant d'un son en des représentations cohérentes est souvent associée à un sentiment d'être envahi par les stimuli extérieurs (Hetrick et al., 2012; McGhie and Chapman, 1961). Afin de mettre en évidence ces modifications perceptuelles, nous avons constitué un corpus de sons incluant des sons de l'environnement pour leur caractère familier et des sons dits abstraits (sons dont il est difficile de reconnaître la source) pour leur caractère bizarre. L'hypothèse sous-jacente reposait sur le fait que le degré de bizarrerie reflèterait la désorganisation perceptuelle des informations auditives. Nous avons ensuite demandé à un groupe de patients schizophrènes et un groupe de sujets témoins d'évaluer le caractère Familier, Bizarre mais aussi Rassurant, Angoissant et Envahissant de ces sons. Les résultats ont montré une tendance significative chez les patients schizophrènes à sous-familiariser les sons familiers tout en percevant l'aspect bizarre des sons bizarres de la même manière que les sujets témoins, en cohérence avec le concept d'ambivalence dans la schizophrénie. En outre, des corrélations avec des données cliniques (échelle PANSS, hallucination, idée délirante, etc.) et électrophysiologiques (composante P50) ont été mis en évidence (Micoulaud-Franchi et al., 2012).

Enfin, le phénomène d'envahissement perceptuel dans la schizophrénie est une des premières altérations perceptuelles décrites par McGhie and Chapman (1961). L'envahissement perceptuel est étayé scientifiquement par des échelles de mesures rétrospectives validées, dont un auto-questionnaire appelé Sensory Gating Inventory (SGI), et l'identification d'un marqueur robuste des anomalies perceptives (le défaut de filtrage sensoriel mesuré par des outils électrophysiologiques, méthode des Potentiels Evoqués). Dans le cadre du travail de master d'Alexandre El-Kaïm supervisé par Jean-Arthur Micoulaud-Franchi, nous avons ainsi mis en place un dispositif d'écoute sonore en temps réel comme outil de mesure de l'envahissement perceptuel dans la schizophrénie en utilisant le synthétiseur de sons d'environnement développé précédemment. Pour cela, nous avons soumis deux populations de sujets (témoins et patients) à l'épreuve de ce dispositif afin de corréler les résultats, aux fins de validation externe de l'outil, avec : 1) l'ou-

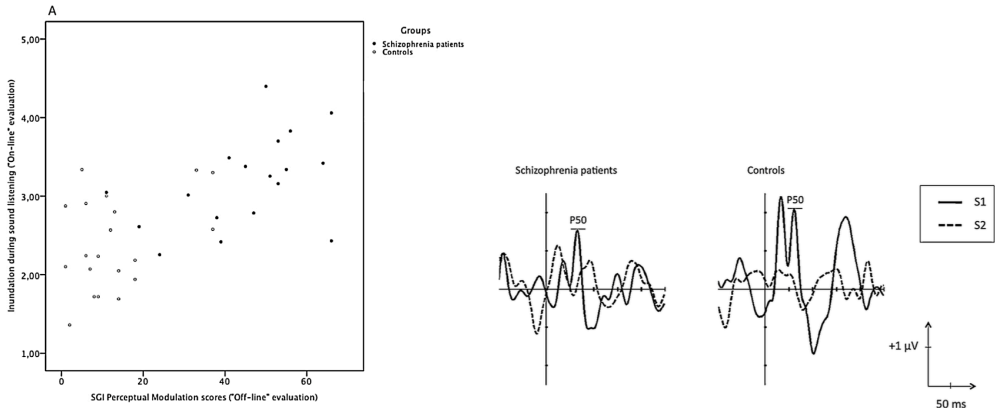


FIGURE 2.9 – **Gauche :** Corrélation significative entre les évaluations de l'envahissement perceptuel mesurées avec le dispositif d'écoute en temps-réel et avec la SGI (pour la dimension "Modulation perceptuelle"). **Droite :** Potentiels P50 révélant un défaut de filtrage sensoriel chez les patients schizphrènes comparé au groupe contrôle. Figures extraites de (El-Kaim et al., 2015).

til actuellement validé en langue française de la SGI (validation psychométrique, Micoulaud-Franchi et al., 2014c) ; 2) l'outil de référence de mesure du filtrage sensoriel par la mesure des potentiels P50 (validation psychophysiologique, Micoulaud-Franchi et al., 2014b). Les résultats obtenus ont permis de valider le dispositif d'écoute en temps réel (Figure 2.9), fondé sur l'utilisation d'un large champ de sons environnementaux et calibrés, et de montrer l'intérêt de la synthèse sonore dans l'étude de la perception dans ce trouble (El-Kaim et al., 2015). A terme, l'ensemble de ces études vise à contribuer à la réalisation d'un test global de diagnostic clinique mais également de remédiation par le son qui pourra être généralisé aux patients atteints d'autres pathologies psychiatriques (Micoulaud-Franchi et al., 2014a).

Chapitre 3

Influencer le comportement par le son

Grâce à l'approche basée sur l'analyse par la synthèse, nous avons mis en évidence un certain nombre d'invariants morphologiques responsables du contenu informatif des sons. Les résultats obtenus et les perspectives entrouvertes par les possibilités de contrôle intuitif des sons ont naturellement mené à explorer un thème de recherche encore peu étudié à ce jour, concernant l'influence manifeste du son sur l'humain. Nous avons cherché à évaluer la manière dont le contrôle de la morphologie des sons permet d'agir de manière implicite sur notre comportement en particulier, en présence d'autres modalités sensorielles associées à la vision ou au système proprioceptif. Grâce à une collaboration étroite et fructueuse avec les chercheurs experts du domaine des sciences cognitives et des sciences du mouvement (Jean-Luc Velay et Jérémy Danna, chercheurs au LNC ; Christophe Bourdin et Lionel Bringoux, enseignants-chercheurs à l'ISM), les interactions son/geste ont été étudiées à mon sens de manière inédite en calibrant les modèles sonores sur la base des connaissances propres aux sciences du vivant. Ces études ont fait notamment l'objet de la thèse d'Etienne Thoret (thèse en co-direction avec S. Ystad). Les applications abordées dans le cadre des projets ANR *Physis*, *MetaSon* et *SoniMove* nous ont conduit à mettre en situation nos outils de synthèse en condition d'utilisation interactive et multimodale.

Le paradigme d'étude utilisé et les résultats mis en évidence dans les expériences conduites ouvrent de nombreuses perspectives pour la compréhension des processus d'intégrations multisensorielles mettant en jeu la vision et l'audition, cette dernière étant souvent considérée comme une modalité secondaire dans l'intégration audiovisuelle du mouvement. De plus dans une perspective écologique, ces résultats offrent un regard nouveau sur la modalité auditive et ses relations avec la motricité, ce qui n'avait jusqu'alors jamais été mis en évidence d'une façon aussi formelle.

3.1 Relation son/mouvement humain

La perception des mouvements biologiques est un sujet d'étude qui date depuis très longtemps. Les travaux de Johansson (1973) ont montré qu'il est possible de reconnaître visuellement un mouvement humain à partir de quelques points lumineux placés sur des jonctions particulières du corps. Dans le cas de gestes graphiques, les études séminales de Viviani et collaborateurs ont montré une relation forte entre la cinématique du mouvement et la trajectoire spatiale, et ont défini une loi en puissance dotée d'un exposant égal à 2/3 reliant la vitesse angulaire du geste et la courbure de la forme dessinée (Lacquaniti et al., 1983; Viviani and Terzuolo, 1982). Cette covariation peut notamment s'expliquer par des contraintes biomécaniques durant la production motrice. De par cette relation, il a été également montré que cette cinématique pouvait perturber la perception visuelle d'une forme induite par le mouvement d'un point lumineux si sa cinématique ne respecte pas la loi en puissance 2/3 (Viviani and Stucchi, 1992). Ainsi, la trajectoire d'un point lumineux parcourant un cercle géométrique sera perçue comme elliptique si ce point subit des accélérations dans les parties verticales ou horizontales (i.e. comme si le point parcourait une ellipse). Ce résultat montre que la perception des formes géométriques est également régie par nos contraintes motrices.

Dans une première expérience, nous avons montré que la loi en puissance 2/3 était également valide pour la modalité auditive. Nous avons en effet mis en évidence le fait qu'il était possible de percevoir un geste graphique humain uniquement par le biais de l'écoute du son de frottement produit (Thoret et al., 2012b,a, 2013c, 2014). Pour cela, nous avons utilisé le modèle de synthèse de sons de frottement précédemment établi afin de contrôler spécifiquement la cinématique du geste via le profil de vitesse. Ce synthétiseur était mis à disposition d'un groupe de sujets à qui nous avons demandé d'ajuster le son pour qu'il évoque un geste humain le plus naturel et fluide possible. Les sujets n'étaient pas informés qu'ils devaient ajuster le coefficient de la loi en puissance. Les résultats ont révélé un ajustement consensuel du coefficient à un valeur qui n'était pas significativement différent de 2/3. De plus, cette expérience a montré que le profil de vitesse permettait de transmettre, via les variations de timbre du son produit, des informations pertinentes sur le geste sous-jacent. En particulier, le son perçu évoquait un geste humain fluide ou uniforme lorsque cet invariant se conformait à la loi en puissance 2/3.

3.1.1 Relation son/forme

Le résultat précédent montre que la cinématique évoquée par le son est fondamentale dans la perception du mouvement. Même si à la base aucun son n'est naturellement associé à une géométrie donnée, la loi en puissance 2/3 suggère un lien intrinsèque entre la dynamique du mouvement et la trajectoire effectuée. Peut-on alors déduire des caractéristiques spatiales d'une dynamique évoquée par un son ? Avec une série de tests de discrimination, nous avons montré

que la perception du profil de vitesse à travers les variations de timbre des sons permet en effet de retrouver dans une certaine mesure la forme dessinée (Thoret et al., 2012b,a, 2013a,c, 2014). En pratique, nous avons constitué un corpus de paires sons/formes qui ont été collectées sur un scripteur dessinant 4 formes simples (cercle, ellipse, arches et trait) sur une tablette graphique. Les sons produits ont été enregistrés avec un microphone placé au-dessus de la tablette. Afin de ne garder que la contribution du profil de vitesse dans le son, les sons enregistrés ont été resynthétisés sous la forme de sons de frottement avec notre synthétiseur uniquement à partir du profil de vitesse mesuré sur la tablette comme paramètre de contrôle. La tâche demandée aux sujets consistait à associer de manière univoque les 4 sons et les 4 formes visuelles présentées aléatoirement sur un écran. Les taux d'association correcte étaient très élevés aussi bien avec les sons naturels qu'avec les sons de synthèse, confirmant ainsi l'importance du profil de vitesse dans l'intégration auditive du mouvement. L'expérience a été reconduite avec un corpus de formes comportant des géométries plus proches (cercle, ellipse, lemniscate et boucle). Comme attendu, les résultats ont mis en évidence des taux d'association correcte moins élevés et une confusion perceptive entre boucle et ellipse, montrant ici les limites de la discrimination perceptive d'une forme évoquée par le son (Thoret et al., 2012b,a, 2013c, 2014). Du point de vue de la théorie écologique de la perception, le profil de vitesse peut être considéré comme un invariant transformationnel permettant de percevoir le geste sous-jacent. Du point de vue de la théorie idéomotrice (Hommel et al., 2001), nous suggérons que le traitement associatif son/forme reposerait sur une représentation amodale des mouvements biologiques basée sur l'intégration des informations sensorielles (visuelle, auditive, sensori-motrice) vers un percept unifié comme proposé par Viviani et al. (1997). Ce point de vue est cohérent avec le fait que des processus communs moteur induisent des contraintes dans la perception des mouvements, et qui se retrouvent aussi bien du point de vue visuel qu'auditif.

Prototype révélé par le son Au sein d'une catégorie de formes géométriques, il a été montré l'existence d'une forme plus particulière que les autres et qui serait la plus représentative de la catégorie considérée. Ces formes dites "prototypiques" ont été mises en évidence du point de vue perceptif (e.g., Rosch, 1973), moteur (e.g., Kalénine et al., 2013) ou haptique (e.g., Theurel et al., 2012). Sur la base de nos résultats précédents, nous avons souhaité montrer qu'une forme prototypique pouvait également être révélée par la modalité auditive. Pour cela, nous avons choisi la catégorie des ellipses et une série d'expériences permettant de comparer les formes révélées à travers différentes modalités (par des reports visuel, moteur et auditif) a été mise en place (Thoret, 2014). Dans l'expérience auditive, les sujets devaient ajuster le son engendré par synthèse afin qu'il évoque la forme elliptique la plus représentative. Dans les expériences visuelles, les sujets devaient ajuster l'excentricité d'une ellipse, présentée en statique et en dynamique (dans ce cas, la trajectoire était déduite du mouvement d'un point lumineux). Enfin, dans

l'expérience motrice, ils devaient dessiner sur une tablette graphique l'ellipse imaginée qui leur semblait la plus représentative. Les résultats obtenus sont remarquables et révèlent l'existence d'un prototype de l'ellipse caractérisée par une excentricité de 0.9 et qui est commun à tous les modes de report considérés, i.e., visuel, moteur et plus particulièrement auditif. L'expérience auditive dans laquelle seule la dynamique était mise en jeu suggère ainsi l'importance de la boucle sensori-motrice dans l'émergence d'un prototype.

3.1.2 Influence du son sur la vision et la motricité

Ces résultats constituent une base de connaissances importantes pour explorer l'impact du son et de ses caractéristiques morphologiques sur le mouvement humain en présence d'autres modalités. Plusieurs études ont montré une influence significative de la modalité auditive sur la perception visuelle du mouvement (Soto-Faraco et al., 2003) et sur la synchronisation sensori-motrice (Repp and Su, 2013). L'expérience que nous avons mis en place a consisté à confronter l'influence des modalités visuelle et auditive dans une tâche de reproduction motrice. Les stimuli audiovisuels ont été engendrés par synthèse afin de contrôler de façon précise les paramètres cinématiques (vitesse, trajectoire). Ils ont été construits en combinant de manière congruente ou incongrue des stimuli visuels et sonores (Figure 3.1, gauche). Les stimuli visuels consistaient en un point lumineux se déplaçant sur un cercle avec une vitesse associée à une cinématique circulaire (mouvement biologique) ou elliptique (mouvement non biologique par rapport à la trajectoire circulaire). Les stimuli sonores étaient des sons de frottement régis par un profil de vitesse associé à une trajectoire circulaire ou elliptique. La tâche demandée aux sujets consistait à reproduire sur une tablette graphique (ils ne voyaient pas leur main) le mouvement du point lumineux. Les résultats ont montré que la reproduction motrice de la trajectoire du point était clairement modifiée par la présence d'un son évoquant une dynamique incongrue (Thoret et al., 2016a). Ainsi la trajectoire circulaire visuelle a été reproduite de façon elliptique en présence d'un son évoquant une trajectoire elliptique même si la dynamique visuelle était cohérente (Figure 3.1, droite).

Ces résultats remarquables démontrent le potentiel du son à influer de manière implicite sur le geste. Ils suggèrent ainsi le développement de nouveaux outils utilisant le son pour l'augmentation ou la substitution d'informations sensorielles, en particulier visuelle ou proprioceptive, dans un contexte de guidage par exemple comme décrit dans le paragraphe suivant.

3.1.3 Guidage par le son

Des travaux fondamentaux menés dans le cadre du projet ANR *MetaSon* et qui se poursuivent dans le projet *SoniMove* ont consisté à définir des stratégies de sonification efficaces dans un contexte de guidage vers une cible. Kramer (1994) définit la sonification comme l'utilisation de

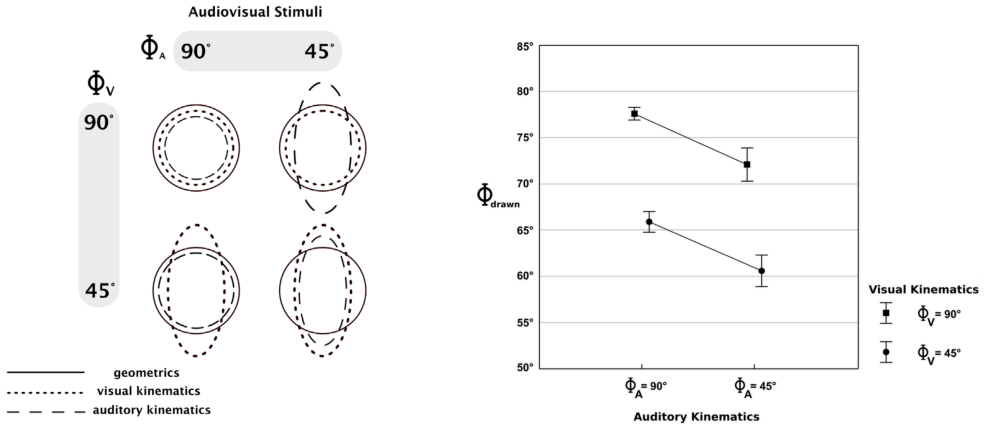


FIGURE 3.1 – **Gauche :** 4 conditions audiovisuelles expérimentales. **Droite :** Phase relative (associée à l'excentricité de l'ellipse reproduite) obtenue dans les 4 conditions expérimentales. Figures extraites de (Thoret et al., 2016a).

sons non verbaux pour transmettre de l'information. A cet égard, le domaine de la sonification a connu une forte expansion depuis quelques années et des applications dans de nombreux domaines peuvent être citées, telles que les interfaces auditives pour l'automobile (Larsson, 2010), les retours auditifs dans le cadre de la physiothérapie (Vogt et al., 2010) et dans le sport d'élite (Schaffert et al., 2010), l'utilisation de représentations auditives sur ordinateur pour aveugles, abordée par le biais d'icônes auditives (Wersényi, 2010), ainsi que la réalité augmentée audio pour la navigation (Mariette, 2010). Cependant, les stratégies de sonification existantes pour le guidage ont souvent été développées sur la base d'attributs spatiaux et dans un contexte spécifique fortement lié à l'application et par conséquent, ne peuvent généralement pas être adaptées directement à d'autres applications.

Stratégies de guidage génériques Dans le cadre des travaux post-doctoraux menés par Gaëtan Parseihian, nous avons ainsi cherché à proposer des stratégies génériques et robustes basées sur des variations de morphologies sonores (et non sur des effets de spatialisation). Nous avons pour cela initié une étude formelle d'évaluation de stratégies sonores basées sur une variation des principaux attributs perceptifs du son tels que la sonie, la hauteur perçue, le tempo, le timbre (inharmonicité, rugosité, brillance, etc.) pour une tâche de guidage vers une cible cachée le long d'un axe unidimensionnel sur une tablette graphique. Cette étude nous a permis d'identifier des morphologies sonores qui favorisent la précision, la rapidité dans l'exécution de la tâche et/ou le non dépassement de la cible (Parseihian et al., 2013, 2015, 2016). En particulier, les stratégies basées sur les modulations temporelles ont permis d'atteindre une meilleure précision, de réduire le nombre d'oscillations autour de la cible et de ne pas la dépasser. Ces stratégies ont en revanche été associées à des temps d'exécution plus longs en comparaison avec des stratégies basées sur la hauteur perçue par exemple. Ces résultats permettent d'établir une première

classification des attributs perceptifs du son en fonction de leur efficacité à informer de leur positionnement par rapport à une cible. Ils conduisent naturellement à de nouveaux protocoles qui sont en cours d'élaboration dans le projet *SoniMove* et dans lequel nous souhaitons étendre le concept de cible en le définissant comme un "état" à atteindre, e.g., une cible dynamique, cibles multiples, suivi de trajectoires, ou bien un geste expert (sportif, musical etc.).

Révéler et remédier le geste d'écriture par le son Une application majeure du projet *MetaSon* a concerné la sonification du geste d'écriture chez les enfants dysgraphiques, application supervisée par Jean-Luc Velay et Jérémy Danna, chercheurs au LNC. L'écriture étant à la base une activité silencieuse, la transformer en une activité bruyante peut paraître curieux en première analyse. Toutefois, justement parce que la modalité auditive est disponible pendant l'écriture et parce qu'elle est mieux à même de permettre la perception de différences fines au plan temporel et dynamique, utiliser des sons pour renseigner sur ce mouvement semble judicieux. Les résultats d'une série d'expériences ont révélé que nous étions capables d'évaluer différemment une écriture d'enfants dysgraphiques par rapport à celle d'enfants ou d'adultes normoscripteurs sur la base seule d'information sonore (Danna et al., 2012a, 2013a,b, 2012b, 2014, 2015b). Un outil de sonification de l'écriture en temps-réel (à partir de données collectées sur une tablette graphique) a été développé sur la base des modèles de synthèse précédemment décrits (chapitre 2) et évalués dans un contexte d'apprentissage de nouveaux caractères (Danna et al., 2015a). L'étape finale de validation de cet outil de sonification est actuellement menée par les chercheurs du LNC auprès d'un groupe d'enfants dysgraphiques et leur thérapeute. Cette validation permettra d'évaluer l'efficacité d'un feedback sonore pour améliorer le geste d'écriture et si cette amélioration persiste dans la durée.

3.1.4 Interaction musicien/instrument

La relation son/mouvement humain est fondamentale dans le contexte du jeu musical. Le musicien ajuste son geste musical en fonction du son produit par l'instrument, mais aurait-il été capable d'apprendre un tel geste expert sans un retour sonore cohérent ? Cette question soulève l'intérêt de la sonification à des fins d'apprentissage d'un geste expert et constitue une des applications majeures du projet *SoniMove*. Dans le cadre des travaux de thèse de Jocelyn Rozé (direction : S. Ystad, PRISM), nous avons étudié l'interaction musicien/instrument dans le cas du violoncelle et l'influence de ses mouvements sur l'interprétation musicale. La performance musicale nécessite la maîtrise des mouvements requis pour produire et moduler le son via l'interaction directe avec l'instrument au niveau de l'archet et du manche. Mais des mouvements dits ancillaires provenant majoritairement de la tête et du torse sont également exécutés par les musiciens experts durant leur performance. Ces mouvements, non directement liés à la production du son, participent à la qualité perçue de la performance générale. En effet, ils semblent

être utiles à la perception émotionnelle du son par l'audience, même en absence de son.

Nous nous sommes ainsi intéressés à la nature de ces comportements posturo-cinétiques en analysant l'influence des mouvements ancillaires sur la qualité du son produit (Rozé et al., 2015a,b; Rozé et al., 2016). Pour cela, nous avons manipulé la contrainte posturale du musicien au cours du jeu afin de comparer des situations de jeu normal à des situations d'immobilisation du torse et/ou de la tête du musicien. Quatre conditions ont été considérées : (1) jeu normal ; (2) immobilisation mentale ; (3) immobilisation physique du torse ; et (4) immobilisation physique du torse et de la tête. Durant le jeu, les mouvements du musicien ont été mesurés grâce à une plateforme de force et un système de capture de mouvements. Les résultats de cette expérience ont montré une influence significative des mouvements posturaux ancillaires dans la qualité du timbre. En particulier, une dégradation du son en terme de "décharnement" a été mise en évidence lorsque le musicien était contraint posturalement. Ces résultats ouvrent des perspectives vers de nouvelles méthodes d'apprentissage qui prendraient davantage en compte le rôle de la posture dans l'acquisition du geste musical expert.

3.2 Le son interactif dans un environnement virtuel

Les mondes virtuels interactifs et les jeux vidéo ont beaucoup évolué au cours des dernières années. L'animation 3D temps réel et le rendu d'images de synthèse réalistes sont maintenant presque devenus un standard. Pourtant, les moteurs actuels de génération sonore se limitent à une lecture d'un corpus de sons enregistrés, impliquant beaucoup de contraintes et de manque de flexibilité (pas de modifications a posteriori) et d'interactivité (pas de traitement d'information en temps-réel). De nombreux responsables de grands studios de post-production s'accordent en effet à dire que l'utilisation des sons enregistrés a atteint sa limite, et ils appellent à la mise en œuvre de nouvelles technologies permettant de générer des sons directement à partir de paramètres physiques, contextuels et sémantiques des objets du monde virtuel. La synthèse des sons constitue ainsi un moyen adapté pour surmonter ces contraintes et ces limites, en créant des moteurs de synthèse innovants permettant des transformations sonores en temps réel, un rendu audio interactif et une adéquation son/image/geste de l'utilisateur. Le projet *Physis* a visé à cette rupture méthodologique en intégrant la synthèse sonore au sein de tels moteurs et en permettant avec ces nouveaux outils une immersion accrue dans les mondes virtuels. Le son pourrait même se substituer, dans certains cas, aux déficiences des autres modalités. Il s'agissait à terme de tenir compte également des différents possibilités d'interactions : par exemple avec les événements, avec d'autres joueurs ou son propre avatar, et nouvelles possibilités de jeu avec les interfaces émergentes.

Afin de proposer des solutions et de montrer le potentiel de la synthèse dans ce contexte d'utilisation, nous avons défini des stratégies de contrôle interactif du synthétiseur directement

à partir des données physiques du monde virtuel (Verron et al., 2013; Pruvost et al., 2015). Le couplage entre un moteur physique/graphique contrôlant le monde virtuel (créé par les logiciels tels que Unity, Blender, UDK) et les modules de synthèse (décrits dans le chapitre 2) a été effectué via le protocole OSC. Nous avons ainsi sélectionné différents cas d'usage mettant le plus à défaut l'approche actuellement adoptée par les designers sonores (cf. Figure 3.2) : a) un jeu interactif de labyrinthe constitué d'obstacles et de surfaces de différents matériaux et formes ; b) scène impliquant un morphing de matériaux, de forme, d'actions ; c) une scène d'éboulement de pierres, permettant de montrer la capacité de gestion d'un très grand nombre d'événements ; d) un jeu interactif de "verre musical" dont le but est de faire chanter un verre en trouvant la bonne combinaison {vitesse-pression} du doigt. Pour ce dernier cas d'usage, nous avons utilisé le modèle de synthèse de sons de friction non linéaire développé par ailleurs (Thoret et al., 2013b). Le contrôle perceptif continu entre frottement, couinement et auto-oscillation, a été développé spécifiquement dans le cadre de ce projet (Pruvost et al., 2015; Thoret et al., 2016b). Par le biais d'une calibration adaptative (évolution du degré de difficulté), ce dernier cas d'usage démontre les possibilités offertes par un tel outil comme aide à l'apprentissage d'un geste particulier.

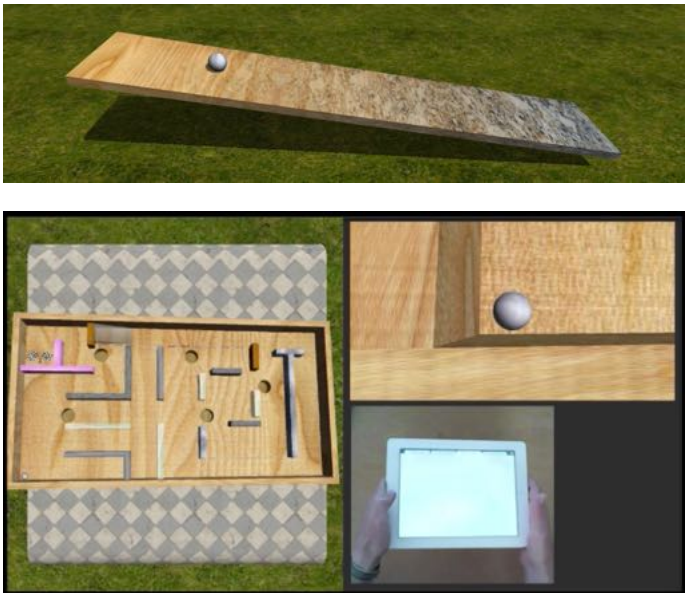


FIGURE 3.2 – **Haut :** Roulement d'une bille sur une planche dont le matériau évolue continûment du bois au métal. Le son est généré par le synthétiseur de sons avec un contrôle continu dans l'espace des matériaux. Figure extraite de (Pruvost et al., 2015). **Bas :** Jeu de labyrinthe dans lequel le son est contrôlé directement à partir des actions de l'utilisateur (via une interface, cf. encadré en bas à droite).

Troisième partie

Liste complète des publications

Articles soumis ou en révision

[1] E. Thoret, M. Aramaki, L. Bringoux, S. Ystad, and R. Kronland-Martinet. When hearing drawing dynamics supports modality-invariant representations of prototypical shapes. *Scientific Report*, en révision.

Articles dans des revues à comité de lecture

[2] B. Bressolette, S. Denjean, V. Roussarie, M. Aramaki, S. Ystad, and R. Kronland-Martinet. Virtual object sonification for gesture controlled human–machine interfaces. *IEEE Consumer Electronics Magazine*, sous presse.

[3] J. Rozé, M. Aramaki, R. Kronland-Martinet, and S. Ystad. Exploring the perceived harshness of cello sounds by morphing and synthesis techniques. *The Journal of the Acoustical Society of America*, 141(3) :2121–2136, 2017.

[4] G. Parseihian, C. Gondre, M. Aramaki, S. Ystad, and R. Kronland-Martinet. Comparison and evaluation of sonification strategies for guidance tasks. *IEEE Transactions on Multimedia*, 18(4) :674–686, 2016.

[5] E. Thoret, M. Aramaki, L. Bringoux, S. Ystad, and R. Kronland-Martinet. Seeing circles and drawing ellipses : when sound biases reproduction of visual motion. *Plos One*, 11(4), 2016.

[6] E. Thoret, M. Aramaki, L. Bringoux, S. Ystad, and R. Kronland-Martinet. When eyes drive hand : influence of non-biological motion on visuo-motor coupling. *Neuroscience Letters*, 612 :225–230, 2016.

[7] E. Thoret, M. Aramaki, C. Gondre, S. Ystad, and R. Kronland-Martinet. Eluding the physical constraints in a nonlinear interaction sound synthesis model for gesture guidance. *Applied Sciences*, 6(7) :192, 2016.

[8] J. Danna, M. Fontaine, V. Paz-Villagrán, C. Gondre, E. Thoret, M. Aramaki, R. Kronland-Martinet, S. Ystad, and J.L. Velay. The effect of real-time auditory feedback on learning new characters. *Human Movement Science*, 43 :216–228, 2015.

[9] J. Danna, V. Paz-Villagrán, C. Gondre, M. Aramaki, R. Kronland-Martinet, S. Ystad, and J.L. Velay. "let me hear your handwriting !" evaluating the movement quality from its sonification. *Plos One*, 10(6) :e0128388, 2015.

[10] A. El-Kaim, M. Aramaki, S. Ystad, R. Kronland-Martinet, M. Cermolacce, J. Naudin, J. Vion-Dury, and J.A. Micoulaud-Franchi. On the correlation between perceptual inundation caused by realistic immersive environmental auditory scenes and the sensory gating inventory in schizophrenia. *European Psychiatry*, 30(5) :606–614, 2015.

[11] J.A. Micoulaud-Franchi, F. Vaillant, R. Lopez, P. Peri, A. Baillif, L. Brandejsky, M.L. Steffen, L. Boyer, R. Richieri, M. Cermolacce, S. Bioulac, M. Aramaki, P. Philip, C. Lancon, and J. Vion-Dury. Sensory gating in adult with attention-deficit/hyperactivity disorder :

Event-evoked potential and perceptual experience reports comparisons with schizophrenia. *Biological Psychology*, 107 :16–23, 2015.

[12] G. Parseihian, S. Ystad, M. Aramaki, and R. Kronland-Martinet. The process of sonification design for guidance tasks. *Audio Mobility*, 9(2), 2015.

[13] S. Conan, O. Derrien, M. Aramaki, S. Ystad, and R. Kronland-Martinet. A synthesis model with intuitive control capabilities for rolling sounds. *IEEE Transactions on Audio, Speech, and Language Processing*, 22(8) :1260–1273, 2014.

[14] S. Conan, E. Thoret, M. Aramaki, O. Derrien, C. Gondre, R. Kronland-Martinet, and S. Ystad. An intuitive synthesizer of continuous interaction sounds : Rubbing, scratching and rolling. *Computer Music Journal*, 38(4) :24–37, 2014.

[15] A. Frey, M. Aramaki, and M. Besson. Conceptual priming for realistic auditory scenes and for auditory words. *Brain and Cognition*, 84(1) :141–152, 2014.

[16] J.A. Micoulaud-Franchi, M. Aramaki, R. Richieri, M. Cermolacce, C. Faget, S. Ystad, R. Kronland-Martinet, C. Lancon, and J. Vion-Dury. Effects of clozapine on perceptual abnormalities and sensory gating. a preliminary cross-sectional study in schizophrenia. *Journal of Clinical Psychopharmacology*, 35(2) :184–187, 2014.

[17] J.A. Micoulaud-Franchi, W.P. Hetrick, M. Aramaki, A. Bolbecker, L. Boyer, S. Ystad, R. Kronland-Martinet, R. Richieri, C. Faget, M. Faugere, A. El-Kaim, M. Cermolacce, C. Lancon, and J. Vion-Dury. Do schizophrenia patients with low p50-suppression report more perceptual anomalies with the sensory gating inventory ? *Schizophrenia Research*, 157(1-3) :157–162, 2014.

[18] J.A. Micoulaud-Franchi, W.P. Hetrick, L. Boyer, A. Bolbecker, M. Aramaki, S. Ystad, R. Richieri, A. El-Kaim, C. Faget, M. Faugere, M. Cermolacce, R. Kronland-Martinet, C. Lancon, and J. Vion-Dury. Validation of the french sensory gating inventory : A confirmatory factor analysis. *Psychiatry Research*, 220(3) :1106–1112, 2014.

[19] E. Thoret, M. Aramaki, R. Kronland-Martinet, J.L. Velay, and S. Ystad. From sound to shape : Auditory perception of drawing movements. *Journal of Experimental Psychology : Human Perception and Performance*, 40 :983–994, 2014.

[20] A. Merer, M. Aramaki, S. Ystad, and R. Kronland-Martinet. Perceptual characterization of motion evoked by sounds for synthesis control purposes. *Association for Computing Machinery, Transactions on Applied Perception*, 10(1) :1–24, 2013.

[21] R. Kronland-Martinet, S. Ystad, and M. Aramaki. High-level control of sound synthesis for sonification processes. *AI & Society, Journal of Knowledge, Culture and Communication*, 27(2) :245–255, 2012.

[22] D. Marelli, M. Aramaki, R. Kronland-Martinet, and C. Verron. An efficient time-frequency method for synthesizing noisy sounds with short transients and narrow spectral components. *IEEE Transactions on Audio, Speech, and Language Processing*, 20(4) :1400–1408, 2012.

[23] J.A. Micoulaud-Franchi, M. Aramaki, A. Merer, M. Cermolacce, S. Ystad, R. Kronland-Martinet, J. Naudin, and J. Vion-Dury. Towards an exploration of feeling of strangeness in schizophrenia : Perspectives on acousmatic and everyday listening. *Journal of Abnormal Psychology*, 121(3) :628–640, 2012.

[24] M. Besson, A. Frey, and M. Aramaki. Is the distinction between intra- and extra-musical meaning implemented in the brain ? : Comment on "Towards a neural basis of processing musical semantics" by Stefan Koelsch. *Physics of Life Reviews*, 8(2) :112–113, June 2011.

[25] M. Aramaki, M. Besson, R. Kronland-Martinet, and S. Ystad. Controlling the perceived material in an impact sound synthesizer. *IEEE Transactions on Audio, Speech, and Language Processing*, 19(2) :301–314, 2011.

[26] J.A. Micoulaud-Franchi, M. Aramaki, A. Merer, M. Cermolacce, S. Ystad, R. Kronland-Martinet, and J. Vion-Dury. Categorization and timbre perception of environmental sounds in schizophrenia. *Psychiatry Research*, 189(1) :149–152, 2011.

[27] M. Aramaki, C. Marie, R. Kronland-Martinet, S. Ystad, and M. Besson. Sound categorization and conceptual priming for nonlinguistic and linguistic sounds. *Journal of Cognitive Neuroscience*, 22(11) :2555–2569, 2010.

[28] D. Marelli, M. Aramaki, R. Kronland-Martinet, and C. Verron. Time-frequency synthesis of noisy sounds with narrow spectral components. *IEEE Transactions on Audio, Speech, and Language Processing*, 18(8) :1929–1940, 2010.

[29] C. Verron, M. Aramaki, R. Kronland-Martinet, and G. Pallone. A 3D immersive synthesizer for environmental sounds. *IEEE Transactions on Audio, Speech, and Language Processing*, 18(6) :1550–1561, 2010.

[30] M. Aramaki, H. Baillères, L. Brancheriau, R. Kronland-Martinet, and S. Ystad. Sound quality assessment of wood for xylophone bars. *Journal of the Acoustical Society of America*, 121(4) :2407–2420, 2007.

[31] C. Magne, C. Astésano, M. Aramaki, S. Ystad, R. Kronland-Martinet, and M. Besson. Influence of syllabic lengthening on semantic processing in spoken french : behavioural and electrophysiological evidence. *Cerebral Cortex*, 17(11) :2659–2668, 2007.

[32] S. Ystad, C. Magne, S. Farner, G. Pallone, M. Aramaki, M. Besson, and R. Kronland-Martinet. Electrophysiological study of algorithmically processed metric/rhythmic variations in language and music. *EURASIP Journal on Audio, Speech, and Music Processing ; Special Issue : "Perceptual Models for Speech, Audio, and Music Processing*", 2007(4), 2007. Article ID 30194.

[33] M. Aramaki and R. Kronland-Martinet. Analysis-synthesis of impact sounds by real-time dynamic filtering. *IEEE Transactions on Audio, Speech, and Language Processing*, 14(2) :695–705, 2006.

[34] M. Aramaki, R. Kronland-Martinet, T. Voinier, and S. Ystad. A percussive sound synthetizer based on physical and perceptual attributes. *Computer Music Journal*, 30(2) :32–41, 2006.

[35] C. Magne, M. Aramaki, C. Astésano, R.L. Gordon, S. Ystad, S. Farner, R. Kronland-Martinet, and M. Besson. Comparison of rythmic processing in language and music : an interdisciplinary approach. *The Journal of Music and Meaning*, Fall 2004/Winter 2005, 2005.

[36] M. Aramaki, J. Bensa, L. Daudet, P. Guillemain, and R. Kronland-Martinet. Resynthesis of coupled piano string vibrations based on physical modeling. *Journal of New Music Research*, 30(3) :213–226, 2001.

[37] M. Aramaki, J. Bensa, L. Daudet, P. Guillemain, and R. Kronland-Martinet. Synthèse des sons musicaux : des modèles mécaniques aux modèles de signaux. *Revue de l'électricité et de l'électronique*, 2 :59–67, 2001.

Articles dans des revues francophones à comité de lecture

[38] J. Danna, V. Paz-Villagran, J.L. Velay, C. Gondre, R. Kronland-Martinet, S. Ystad, M. Aramaki, E. Thoret, S. Conan, T. Voinier, H. Omer, and B. Torrésani. Sonifier l'écriture : un outil pour le diagnostic et la remédiation de la dysgraphie. *Développements*, 12 :32–40, 2012.

Communications avec actes de congrès

[39] T. Bordonné, M. Dias-Alves, M. Aramaki, S. Ystad, and R. Kronland-Martinet. Assessing sound perception through vocal imitations of sounds that evoke movements and materials. In *Proceedings of the 13th International Symposium on Computer Music Multidisciplinary Research (CMMR) "Music Technology with Swing"*, pages 272–282, Matosinhos, Portugal, 2017.

[40] G. Parseihian, M. Aramaki, S. Ystad, and R. Kronland-Martinet. Sonification strategies for dynamic guidance tasks : example with a driving game. In *Proceedings of the 13th International Symposium on Computer Music Multidisciplinary Research (CMMR) "Music Technology with Swing"*, pages 283–294, Matosinhos, Portugal, 2017.

[41] J. Rozé, R. Kronland-Martinet, M. Aramaki, C. Bourdin, and S. Ystad. Investigating the effects of a postural constraint on the cellists' bowing movement and timbral quality. In *Proceedings of the 12th International Symposium on Computer Music Multidisciplinary Research (CMMR) "Bridging People and Sound"*, pages 83–98, São Paolo, Brésil, 2016.

[42] L. Pruvost, B. Scherrer, M. Aramaki, S. Ystad, and R. Kronland-Martinet. Perception-based interactive sound synthesis of morphing solids' interactions. In *Proceedings of the Siggraph Asia 2015*, Kobe, Japon, 2-5 Novembre 2015 2015.

[43] J. Rozé, M. Aramaki, C. Bourdin, D. Chadefaux, M. Dufrenne, R. Kronland-Martinet, T. Voinier, and S. Ystad. Exploring the influence of cellists' postural movements on musical expressivity. In *Proceedings of the International Conference on the Multimodal Experience of Music*, Sheffield, UK, 2015.

[44] J. Rozé, M. Aramaki, R. Kronland-Martinet, T. Voinier, C. Bourdin, D. Chadefaux, M. Dufrenne, and S. Ystad. Assessing the influence of constraints on cellists' postural displacements and musical expressivity. In *Proceedings of the 11th International Symposium on Computer Music Multidisciplinary Research (CMMR) "Music, Mind and Embodiment"*, Plymouth, UK, 16-19 Juin 2015 2015.

[45] L. Bosc, E. Diaz, V. Roussarie, M. Aramaki, J.L. Velay, R. Kronland-Martinet, J.C. Charmard, and J. Couturier. Rôle du contenu évocatoire des sons dans l'efficacité de la sonification extérieure de véhicules silencieux. In *Actes du 12ème Congrès Français d'Acoustique*, pages 657–663, Poitiers, France, 22-25 Avril 2014.

[46] S. Conan, E. Thoret, M. Aramaki, O. Derrien, C. Gondre, R. Kronland-Martinet, and S. Ystad. Navigating in a space of synthesized interaction-sounds : rubbing, scratching and rolling sounds. In *Proceedings of the 16th International Conference on Digital Audio Effects (DAFx-13)*, Maynooth, Ireland, 2013.

[47] S. Conan, E. Thoret, C. Gondre, M. Aramaki, R. Kronland-Martinet, and S. Ystad. An intuitive synthesizer of sustained interaction sounds. In *Proceedings of the 10th International*

Symposium on Computer Music Multidisciplinary Research (CMMR 2013) "Sound, Music and Motion", Marseille, France, 2013.

[48] J. Danna, J.L. Velay, V. Paz-Villagrán, A. Capel, C. Petroz, C. Gondre, E. Thoret, M. Aramaki, S. Ystad, and R. Kronland-Martinet. Handwriting movement sonification for the rehabilitation of dysgraphia. In *Proceedings of the 10th International Symposium on Computer Music Multidisciplinary Research (CMMR 2013) "Sound, Music and Motion"*, 2013.

[49] G. Parseihian, C. Gondre, M. Aramaki, R. Kronland-Martinet, and S. Ystad. Exploring the usability of sound strategies for guiding task : toward a generalization of sonification design. In *Proceedings of the 10th International Symposium on Computer Music Multidisciplinary Research (CMMR 2013) "Sound, Music and Motion"*, Marseille, France, 2013.

[50] C. Rakovec, M. Aramaki, and R. Kronland-Martinet. Perception of material and shape of impacted everyday objects. In *Proceedings of the 10th International Symposium on Computer Music Multidisciplinary Research (CMMR 2013) "Sound, Music and Motion"*, Marseille, France, 2013.

[51] E. Thoret, M. Aramaki, C. Bourdin, L. Bringoux, R. Kronland-Martinet, and S. Ystad. Synchronizing gestures with friction sounds : Work in progress (conférence invitée). In *Proceedings of the 10th International Symposium on Computer Music Multidisciplinary Research (CMMR 2013) "Sound, Music and Motion"*, Marseille, France, 2013.

[52] E. Thoret, M. Aramaki, C. Gondre, R. Kronland-Martinet, and S. Ystad. Controlling a non linear friction model for evocative sound synthesis applications. In *Proceedings of the 16th International Conference on Digital Audio Effects (DAFx-13)*, Maynooth, Ireland, 2013.

[53] C. Verron, M. Aramaki, A. Gonot, T. Scotti, C. Rakovec, A. Mingasson, and R. Kronland-Martinet. Event-driven interactive solid sound synthesis. In *Proceedings of the 10th International Symposium on Computer Music Multidisciplinary Research (CMMR 2013) "Sound, Music and Motion"*, pages 960–966, Marseille, France, 2013.

[54] S. Conan, M. Aramaki, R. Kronland-Martinet, E. Thoret, and S. Ystad. Perceptual differences between sounds produced by different continuous interactions. In *Proceedings of the 11th Congrès Français d'Acoustique*, pages 409–414, Nantes, France, 23-27 Avril 2012.

[55] S. Conan, M. Aramaki, R. Kronland-Martinet, and S. Ystad. Rolling sound synthesis : Work in progress. In *Proceedings of the 9th International Symposium on Computer Music Modelling and Retrieval (CMMR 2012) - Music & Emotions*, pages 257–264, Londres, Royaume-Uni, 19-22 Juin 2012.

[56] A. Sirdey, O. Derrien, R. Kronland-Martinet, and M. Aramaki. ESPRIT in Gabor frames. In *Proceedings of the 45th Conference of Audio Engineering Society*, pages 305–313, Helsinki, Finlande, 1-4 Mars 2012.

[57] E. Thoret, M. Aramaki, R. Kronland-Martinet, J.L. Velay, and S. Ystad. From shape to sound : Sonification of two dimensional curves by reenaction of biological movements. In *Proceedings of the 9th International Symposium on Computer Music Modelling and Retrieval (CMMR 2012) - Music & Emotions*, pages 278–285, Londres, Royaume-Uni, 19-22 Juin 2012.

[58] E. Thoret, M. Aramaki, R. Kronland-Martinet, J.L. Velay, and S. Ystad. Sonifying drawings : characterization of perceptual attributes of sounds produced by human gestures. In *Proceedings of the 11th Congrès Français d'Acoustique*, pages 1089–1094, Nantes, France, 23-27 Avril 2012.

[59] A. Sirdey, O. Derrien, R. Kronland-Martinet, and M. Aramaki. Modal analysis of impact sounds with ESPRIT in Gabor transforms. In *Proceedings of the Digital Audio Effects Conference : DAFx11*, pages 387–392, Paris, France, 19-23 Septembre 2011.

[60] M. Aramaki and M. Besson. Approche électrophysiologique de la sémiotique des sons (conférence invitée). In *Musique & Sciences Cognitives, Actes de la journée du 21 Mars 2007, 3ièmes Rencontres Interartistiques de l'Observatoire Musical Français*, Conférences et Séminaires, N° 46, pages 99–107, 2010.

[61] D. Marelli, M. Aramaki, R. Kronland-Martinet, and C. Verron. A unified time-frequency method for synthesizing noisy sounds with short transients and narrow spectral components. In *Proceedings of the 35th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 29–32, Dallas, Etats-Unis, 14-19 Mars 2010. IEEE.

[62] A. Merer, S. Ystad, R. Kronland-Martinet, and M. Aramaki. On the potentiality of abstract sounds in perception research. In *Proceedings of the 7th International Symposium on Computer Music Modeling and Retrieval (CMMR 2010) - Music is in the sound*, pages 207–219, Málaga, Espagne, 21-24 Juin 2010.

[63] C. Verron, M. Aramaki, R. Kronland-Martinet, and G. Pallone. Contrôle intuitif d'un synthétiseur d'environnements sonores spatialisés. In *Proceedings of the 10th Congrès Français d'Acoustique*, Lyon, France, 12-16 Avril 2010.

[64] M. Aramaki, C. Gondre, R. Kronland-Martinet, T. Voinier, and S. Ystad. Thinking the sounds : an intuitive control of an impact sound synthesizer. In M. Aramaki, R. Kronland-Martinet, S. Ystad, and K. Jensen, editors, *Proceedings of the 15th International Conference on Auditory Display (ICAD 2009)*, pages 119–124, Copenhague, Danemark, 18-21 Mai 2009. Re :New - Digital Arts Forum.

[65] C. Verron, M. Aramaki, R. Kronland-Martinet, and G. Pallone. Analysis/synthesis and spatialization of noisy environmental sounds (conférence invitée). In M. Aramaki, R. Kronland-Martinet, S. Ystad, and K. Jensen, editors, *Proceedings of the 15th International Conference on Auditory Display (ICAD 2009)*, pages 36–40, Copenhague, Danemark, 18-21 Mai 2009. Re :New - Digital Arts Forum.

[66] C. Verron, G. Pallone, M. Aramaki, and R. Kronland-Martinet. Controlling a spatialized environmental sound synthesizer. In *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 321–324, New Paltz, NY, 18-21 Octobre 2009.

[67] M. Aramaki, L. Brancheriau, R. Kronland-Martinet, and S. Ystad. Perception of impacted materials : sound retrieval and synthesis control perspectives. In K. Jensen, editor, *Proceedings of the Computer in Music Modeling and Retrieval - Genesis of Meaning in Digital Art*, pages 1–8, Copenhague, Danemark, 19-23 Mai 2008. Re :New - Digital Arts Forum.

[68] C. Verron, M. Aramaki, R. Kronland-Martinet, and G. Pallone. Spatialized additive synthesis of environmental sounds. In *Proceedings of the 125th Convention of Audio Engineering Society*, San Francisco, USA, 2-6 Octobre 2008. Convention Paper 7509.

[69] M. Aramaki, R. Kronland-Martinet, T. Voinier, and S. Ystad. Timbre control of a real-time percussive synthesizer (conférence invitée). In A. Calvo-Manzano, A. Pérez-López, and S. Santiago, editors, *Proceedings of the 19th International Congress on Acoustics*, Madrid, Espagne, 2-7 Septembre 2007.

[70] A Merer, S. Ystad, R. Kronland-Martinet, M. Aramaki, M. Besson, and J.L. Velay. Perceptual categorization of moving sounds for synthesis application. In *Proceedings of the International Computer Music Conference (ICMC'07) - Immersed Music*, pages 69–72, Copenhague, Danemark, 27-31 Août 2007.

[71] C. Verron, R. Kronland-Martinet, G. Pallone, and M. Aramaki. A spatialized additive synthesizer. In E. Schubert, K. Buckley, R. Eliott, B. Koboroff, J. Chen, and C. Stevens, editors, *Proceedings of the inaugural International Conference Music Communication Science (ICoMCS)*, pages 167–171, Sydney, Australie, 5-7 Décembre 2007.

[72] M. Aramaki, L. Brancheriau, H. Baillères, R. Kronland-Martinet, and S. Ystad. Relationship between sound classification of xylophone-like bars and wood species properties (conférence invitée). In J. Eberhardsteiner, H.A. Mang, and H. Waubke, editors, *Proceedings of the 13th International Congress on Sound and Vibration (ICSV13)*, Vienne, Autriche, 2-6 juillet 2006.

[73] M. Aramaki, R. Kronland-Martinet, T. Voinier, and S. Ystad. Synthesis and perceptual manipulation of percussive sounds. In *Proceedings of the International Computer Music Conference (ICMC 2005) - Free Sound*, pages 335–338, Barcelone, Espagne, 4-10 Septembre 2005.

[74] M. Aramaki and R. Kronland-Martinet. Analysis-synthesis of impact sounds. In *Proceedings of the 12th European Signal Processing Conference (EUSIPCO-2004)*, pages 1769–1772, Vienne, Autriche, 6-10 Septembre 2004.

[75] M. Aramaki, J. Bensa, L. Daudet, P. Guillemain, and R. Kronland-Martinet. Synthesis of musical sounds : from mechanical models to signal processing models. In *Proceedings of the 2nd Symposium International PSIP'2001 (Physics in Signal and Image Processing)*, pages 231–236, 23-24 Janvier 2001.

[76] M. Aramaki and O. Gipouloux. Synthesis of bi-dimensional structure vibrations using digital waveguides adapted by optimization technics. In *Proceedings of the Workshop on Current Research Directions in Computer Music - Mosart project "Music Orchestration Systems in Algorithmic Research and Technology*", pages 119–123, Barcelone, Espagne, 15-17 Novembre 2001.

[77] M. Aramaki and P. Guillemain. Synthesis of membrane and thin plate vibrations using adapted digital waveguides. In *Proceedings of the International Symposium of Musical Acoustics*, pages 225–228, Pérouse, Italie, 10-14 Septembre 2001.

[78] M. Aramaki, P. Guillemain, N. Hoeb, and A. Roure. From piano string vibrations to the acoustic field radiated by the soundboard. In *Proceedings of the International Computer Music Conference (ICMC 2001)*, volume 2001, pages 111–114, La Havane, Cuba, 15-17 Septembre 2001.

[79] M. Aramaki, J. Bensa, L. Daudet, P. Guillemain, and R. Kronland-Martinet. Synthesis model of piano strings vibrations using coupled digital waveguides. In *Actes du 5ème Congrès Français d'Acoustique*, pages 277–280, Lausanne, Suisse, 3-7 Septembre 2000.

[80] M. Aramaki, L. Brancheriau, H. Baillères, P. Guillemain, R. Kronland-Martinet, and J.D. Lanvin. Caractérisation des produits collés à base de bois par l'analyse des vibrations dans le domaine acoustique. In *Actes du 5ème Congrès des Sciences et Industries du bois*, Nantes, France, 23-25 Mai 2000.

[81] M. Aramaki, L. Brancheriau, H. Baillères, P. Guillemain, R. Kronland-Martinet, and J.D. Lanvin. Modeling vibrations response of wooden beams with singularities using mechanical models. In *Proceedings of the 12th International Symposium on Non-Destructive Testing of Wood*, page 458, Sopron, Hongrie, 13-15 Septembre 2000.

[82] M. Aramaki, L. Brancheriau, H. Baillères, P. Guillemain, R. Kronland-Martinet, and J.D. Lanvin. Modeling vibrations response of wooden beams with singularities using sound synthesis models. In *Proceedings of the 12th International Symposium on Non-Destructive Testing of Wood*, pages 113–121, Sopron, Hongrie, 13-15 Septembre 2000.

[83] M. Aramaki, J. Bensa, L. Daudet, R. Kronland-Martinet, and P. Guillemain. Resynthesis of coupled piano string vibrations based on physical modeling. In *Proceedings of the Digital Audio Effects Conference : DAFx99*, pages 135–138, Trondheim, Norvège, 9-11 Décembre 1999.

Communications sans actes de congrès

[84] G. Parseihian, C. Gondre, M. Aramaki, R. Kronland-Martinet, and S. Ystad. Influence of sound morphologies on movement behavior in sound-guided tasks. In *15th International Multisensory Research Forum (IMRF)*, Amsterdam, Hollande, 2014.

[85] E. Thoret, M. Aramaki, L. Bringoux, R. Kronland-Martinet, and S. Ystad. When acoustic stimuli turn visual circles into ellipses : sounds evoking accelerations modify visuo-motor coupling. In *15th International Multisensory Research Forum (IMRF)*, Amsterdam, Hollande, 2014.

[86] E. Thoret, M. Aramaki, R. Kronland-Martinet, J.L. Velay, and S. Ystad. Auditory perception of biological movements : an evidence of cognitive specifities from sound synthesis. In *15th International Multisensory Research Forum (IMRF)*, Amsterdam, Hollande, 2014.

[87] M. Aramaki, R. Kronland-Martinet, and S. Ystad. Intuitive control for environmental sound synthesis. In *8th International Symposium on Computer Music Modelling and Retrieval (CMMR 2011) joint with 20th International Symposium on Frontiers of Research on Speech and Music (FRSM 2011)*, Bhubaneswar, Inde, 9-12 Mars 2011.

[88] A. Frey, M. Aramaki, and M. Besson. Do auditory environmental scenes have meaning ? In *XI International Conference on Cognitive Neuroscience*, Palma, Mallorque, Espagne, 25-29 Septembre 2011.

[89] J.A. Micoulaud-Franchi, M. Aramaki, A. Merer, M. Cermolacce, S. Ystad, R. Kronland-Martinet, and J. Vion-Dury. Perception of the impacted sound prosody in schizophrenia. In *10ème Colloque de la Société des Neurosciences*, Marseille, France, 24-27 Mai 2011.

[90] J.A. Micoulaud-Franchi, M. Aramaki, A. Merer, M. Cermolacce, S. Ystad, R. Kronland-Martinet, L. Boyer, and J. Vion-Dury. Reconnaissance du bizarre et du familier dans la perception de sons inouïs chez le patient schizophrène : Le problème de la catégorisation. In *Actes du 8ème Congrès de l'Encéphale*, Paris, France, 21-23 Janvier 2010.

[91] J.A. Micoulaud-Franchi, M. Aramaki, A. Merer, M. Cermolacce, S. Ystad, R. Kronland-Martinet, L. Boyer, and J. Vion-Dury. Reconnaissance du bizarre et du familier dans la perception de sons inouïs chez le patient schizophrène. In *1er Congrès Français de Psychiatrie*, Nice, France, 2-5 Décembre 2009.

[92] A. Merer, M. Aramaki, R. Kronland-Martinet, and S. Ystad. Toward synthesis tools using “evocation" as control parameters. In *Acoustics'08*, Paris, France, 29 Juin-4 Juillet 2008.

[93] C. Verron, M. Aramaki, R. Kronland-Martinet, and G. Pallone. Spatialized additive synthesis. In *Acoustics'08*, Paris, France, 29 Juin-4 Juillet 2008.

[94] M. Aramaki, R. Kronland-Martinet, S. Ystad, and M. Besson. Catégorisation sonore de matériaux frappés : approches perceptive et cognitive. In *Journées fondatrices Perception Sonore*, INSA Lyon, France, 18-19 Janvier 2007.

Communications dans des conférences nationales

[95] J. Danna, M. Fontaine, C. Gondre, E. Thoret, M. Aramaki, R. Kronland-Martinet, S. Ystad, and J.L. Velay. Sonifier le geste d'écriture peut-il faciliter son apprentissage ? In *Actes du 15ème Congrès International de l'ACAPS*, Grenoble, France, 2013.

[96] J. Danna, V. Paz-Villagrán, C. Gondre, M. Aramaki, R. Kronland-Martinet, S. Ystad, and J.L. Velay. « ecoute cette écriture ! » sonifier le mouvement d'écriture pour le diagnostic de la dysgraphie. In *2ème Journées Perception Sonore (JPS)*, Marseille, France, 2012.

[97] J. Danna, V. Paz-Villagrán, J.L. Velay, C. Gondre, R. Kronland-Martinet, S. Ystad, M. Aramaki, E. Thoret, S. Conan, T. Voinier, H. Omer, and B. Torrésani. Sonifier l'écriture : un outil pour le diagnostic et la remédiation de la dysgraphie. In *IIIèmes Journée Scientifique du Centre de Référence des Troubles d'Apprentissage (CERTA)/RESODYS*, Marseille, France, 2012.

Chapitres d'ouvrage

[98] J. Rozé, R. Kronland-Martinet, M. Aramaki, C. Bourdin, and S. Ystad. Eluding the influence of postural constraints on cellists' bowing movements and timbral quality. In *Bridging People and Sound, Post-proceedings of 12th International Symposium on Computer Music Modeling and Retrieval (CMMR 2016)*, pages 109–124. Springer International Publishing, 2017.

[99] J. Rozé, M. Aramaki, R. Kronland-Martinet, T. Voinier, C. Bourdin, D. Chadefaux, M. Dufrenne, and S. Ystad. Exploring the effects of constraints on the cellist's postural displacements and their musical expressivity. In *Music, Mind and Embodiment, Post-proceedings of 11th International Symposium on Computer Music Modeling and Retrieval (CMMR 2015)*, pages 22–41. Springer International Publishing, 2016.

[100] M. Aramaki, J.A. Micoulaud-Franchi, R. Kronland-Martinet, J. Vion-Dury, and S. Ystad. Prospective view on sound synthesis bci control in light of two paradigms of cognitive neuroscience. In Eduardo Reck Miranda and Julien Castet, editors, *Guide to Brain-Computer Music Interfacing*, pages 61 – 87. Springer Berlin Heidelberg, 2014.

[101] J. Danna, J.L. Velay, V. Paz-Villagrán, A. Capel, C. Petroz, C. Gondre, E. Thoret, M. Aramaki, S. Ystad, and R. Kronland-Martinet. Movement sonification for the diagnosis and the rehabilitation of graphomotor disorders. In *Sound, Music and Motion, Post-proceedings of 10th International Symposium on Computer Music Modeling and Retrieval (CMMR 2013)*, pages 246–255. Springer International Publishing, 2014.

[102] E. Thoret, M. Aramaki, L. Bringoux, R. Kronland-Martinet, and S. Ystad. Audio-motor synchronization : the effect of mapping between kinematics and acoustic cues on geometric motor features. In *Sound, Music and Motion, Post-proceedings of 10th International Symposium on Computer Music Modeling and Retrieval (CMMR 2013)*, pages 234–245. Springer International Publishing, 2014.

[103] S. Conan, M. Aramaki, R. Kronland-Martinet, and S. Ystad. Intuitive control of rolling sound synthesis. In *From Sounds to Music and Emotions, Post-proceedings of 9th International Symposium on Computer Music Modeling and Retrieval (CMMR 2012)*, volume 7900 of *Lecture Notes in Computer Science*, pages 99–109. Springer-Verlag Berlin Heidelberg, 2013.

[104] E. Thoret, M. Aramaki, R. Kronland-Martinet, and S. Ystad. Reenacting sensorimotor features of drawing movements from friction sounds. In *From Sounds to Music and Emotions, Post-proceedings of 9th International Symposium on Computer Music Modeling and Retrieval (CMMR 2012)*, number 7900 in Lecture Notes in Computer Science, pages 130–153. Springer-Verlag Berlin Heidelberg, 2013.

[105] M. Aramaki, R. Kronland-Martinet, and S. Ystad. Perceptual control of environmental sound synthesis. In S. Ystad, M. Aramaki, R. Kronland-Martinet, K. Jensen, and S. Mohanty, editors, *Speech, Sound and Music Processing : Embracing Research in India*, volume 7172 of *Lecture Notes in Computer Science*, pages 172–186. Springer-Verlag Berlin Heidelberg, 2012.

[106] A. Merer, S. Ystad, R. Kronland-Martinet, and M. Aramaki. Abstract sounds and their applications in audio and perception research. In S. Ystad, M. Aramaki, R. Kronland-Martinet, and K. Jensen, editors, *Exploring Music Contents*, volume 6684 of *Lecture Notes in Computer Science*, pages 176–187. Springer Berlin Heidelberg, 2011.

[107] M. Aramaki, C. Gondre, R. Kronland-Martinet, T. Voinier, and S. Ystad. Imagine the sounds : an intuitive control of an impact sound synthesizer. In S. Ystad, M. Aramaki, R. Kronland-Martinet, and K. Jensen, editors, *Auditory Display*, volume 5954 of *Lecture Notes in Computer Science*, pages 408–421. Springer Berlin Heidelberg, 2010.

[108] M. Besson, D. Schön, M. Aramaki, C. Marie, and A. Frey. La sémiotique des sons : approches acoustique et électrophysiologique. In Mario Borillo, editor, *Dans l'Atelier de l'Art : Expériences Cognitives*, pages 155–169. Champ Vallon, 2010.

[109] C. Verron, M. Aramaki, R. Kronland-Martinet, and G. Pallone. Spatialized synthesis of noisy environmental sounds. In S. Ystad, M. Aramaki, R. Kronland-Martinet, and K. Jensen, editors, *Auditory Display*, volume 5954 of *Lecture Notes in Computer Science*, pages 392–407. Springer Berlin Heidelberg, 2010.

[110] M. Aramaki, M. Besson, R. Kronland-Martinet, and S. Ystad. Timbre perception of sounds from impacted materials : behavioral, electrophysiological and acoustic approaches. In S. Ystad, R. Kronland-Martinet, and K. Jensen, editors, *Computer Music Modeling and Retrieval - Genesis of Meaning of Sound and Music*, volume 5493 of *LNCS*, pages 1–17. Springer Berlin Heidelberg, 2009.

[111] M. Aramaki, L. Brancheriau, R. Kronland-Martinet, and S. Ystad. Perception of impacted materials : sound retrieval and synthesis control perspectives. In S. Ystad, R. Kronland-Martinet, and K. Jensen, editors, *Computer Music Modeling and Retrieval - Genesis of Meaning of Sound and Music*, volume 5493 of *LNCS*, pages 134–146. Springer Berlin Heidelberg, 2009.

[112] M. Aramaki, J. Vion-Dury, D. Schön, C. Marie, and M. Besson. Une approche interdisciplinaire de la sémiotique des sons. In *Le "Nouveau" en Musique et en Musicologie : Esthétiques Technologies Langages*, pages 125–135. Libreria Musicale Italiana, 2008.

[113] A. Merer, S. Ystad, R. Kronland-Martinet, and M. Aramaki. Semiotics of sounds evoking motions : Categorization and acoustic features. In *Computer Music Modeling and Retrieval - Sense of Sounds*, volume 4969 of *Lecture Notes in Computer Science*, pages 139–158. Springer Berlin Heidelberg, Copenhague, Danemark, 2008.

Editions d'ouvrage

[114] M. Aramaki, R. Kronland-Martinet, and S. Ystad, editors. *Bridging People and Sound*, volume 10525 of *Lecture Notes in Computer Science*. Springer International Publishing, 2017.

[115] R. Kronland-Martinet, M. Aramaki, and S. Ystad, editors. *Music, Mind and Embodiment*, volume 9617 of *Lecture Notes in Computer Science*. Springer International Publishing, 2016.

[116] M. Aramaki, O. Derrien, R. Kronland-Martinet, and S. Ystad, editors. *Sound, Music and Motion*, volume 8905 of *Lecture Notes in Computer Science*. Springer International Publishing, 2014.

[117] M. Aramaki, M. Barthet, R. Kronland-Martinet, and S. Ystad, editors. *From Sounds to Music and Emotions*, volume 7900 of *Lecture Notes in Computer Science*. Springer-Verlag Berlin Heidelberg, 2013.

[118] S. Ystad, M. Aramaki, R. Kronland-Martinet, K. Jensen, and S. Mohanty, editors. *Speech, Sound and Music Processing : Embracing Research in India*, volume 7172 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2012.

[119] S. Ystad, M. Aramaki, R. Kronland-Martinet, and K. Jensen, editors. *Exploring Music Contents*, volume 6684 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2011.

[120] S. Ystad, M. Aramaki, R. Kronland-Martinet, and K. Jensen, editors. *Auditory Display*, volume 5954 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2010.

Quatrième partie

Activités liées à la recherche

Cette partie du document décrit les activités connexes à celles de mes recherches. Elles sont organisées selon mes activités d'enseignement, d'encadrement, d'organisation de congrès ainsi que mes participations aux projets collaboratifs notamment financés par l'Agence Nationale de la Recherche (ANR).

Enseignement, formation

Activités d'enseignement

- De 2002 à 2004, j'ai assuré des travaux dirigés et travaux pratiques de Traitement de Signal (Matlab, Electronique) au département "Génie des Télécommunications et Réseaux" de l'IUT d'Aix-en-Provence (site de Luminy).
- A partir de 2008, mes activités d'enseignement se situent dans le Master Recherche Mention "Mécanique, Physique et Ingénierie (MPI)", Spécialité Acoustique de Aix-Marseille Université / Ecole Centrale Marseille.
  Site : http://www.master-acoustique.lma.cnrs-mrs.fr/
  Depuis 2012, je fais partie de l'équipe de direction et je suis responsable de l'organisation et du suivi des stages du Master 2. J'enseigne également dans le module "Analyse, synthèse et transformation des sons" (responsable : S. Ystad) avec une série de cours que j'ai construit sur les thématiques proches de mes recherches, notamment les processus de synthèse et du contrôle du son numérique ainsi qu'une introduction à la perception des sons de l'environnement et à la sémiotique des sons.
- J'assure depuis 2015 un cours appliqué en Master 1 MPI sur une initiation au son numérique (échantillonnage, filtrage, quantification, représentations temporelle et fréquentielle) et à la synthèse sonore (synthèse additive).
- Dans le cadre de la nouvelle offre de formation pour la rentrée 2018, un projet de création d'une mention de Master "Son, Musique et Sciences" à Aix-Marseille Université, en lien direct avec la création du laboratoire PRISM est en cours d'élaboration. Cette mention vise à offrir une formation transdisciplinaire, entre théorie et pratiques, autour des thématiques relatives aux sciences de la musique et du son. Elle sera organisée selon 2 parcours, l'un axé sur "Musicologie et création" (rattachée à l'UFR ALLSH) et l'autre dont j'assurerai la responsabilité sur "Ingénierie et conception sonore" (rattachée à l'UFR Sciences).

Activités d'encadrement

Depuis 2006, j'ai participé à l'encadrement de nombreux stagiaires de niveau licence, master 1 et master 2. J'ai également eu l'opportunité d'assurer la co-direction de 6 doctorants dont 3 sont en cours :

Antoine BOURACHOT (Octobre 2016 – en cours) : Financement MENRT

Co-direction avec R. Kronland-Martinet (PRISM)

Titre : "Nouvelles stratégies de contrôle perceptif des processus de synthèse sonore"

Résumé : Le travail de thèse envisagé a pour objectif d'aborder un des grands défis actuels dans ce domaine et qui concernent les problématiques de contrôle intuitif des sons par synthèse. L'hypothèse sera basée sur l'existence d'invariants morphologiques sonores responsables des évocations induites et définis à partir de considérations physiques et de caractérisations acoustiques des signaux produits. Ces travaux ouvrent la porte à la constitution d'un véritable langage des sons dont le contenu informatif sera basé sur la notion de métaphores sonores. Enfin, il s'agira de mettre en œuvre de stratégies de navigation dans des espaces sonores en cohérence avec la perception et le ressenti dans un contexte multimodal large.

Manuel DIAS ALVEZ (Octobre 2016 – en cours) : Psychiatre

Co-direction avec J. Vion-Dury (LNC)

Titre : "Perception et reproduction des invariants cinématiques et acoustiques au travers de sons de gestes sonifiés"

Résumé : Les mouvements humains sont régis par des lois cinématiques. Parmi elles, la loi de puissance 1/3 traduit une relation spécifique entre le rayon de courbure et la vitesse du mouvement lors d'un geste de dessin. Des travaux ont récemment suggéré que les profils de vitesse respectant cette loi sont perceptibles et pouvaient être considérés comme un invariant acoustique transformationnel. Une première étude s'est intéressée à la perception et la reproduction de ces invariants au travers de gestes sonifiés grâce à un procédé de synthèse sonore validé. Les premiers résultats démontrent que les sons des gestes sonifiés respectant la loi de puissance sont mieux reproduits que ceux ne la respectant pas et confirment l'intégration perceptive auditive de la loi de puissance mais aussi l'importance des contraintes dynamiques dans lesquelles sont réalisées les gestes. Le projet de thèse s'attachera à continuer cette étude d'invariants cinématiques et acoustiques associés au geste humain. Des analyses adaptées permettront de préciser la perception des invariants cinématiques et acoustiques par la restitution motrice et fourniront des connaissances sur la perception auditive des invariants.

Thomas BORDONNE (Novembre 2015 – en cours) : Financement ANR

Co-direction avec R. Kronland-Martinet (PRISM)

Titre : “Contrôle intuitif de la synthèse sonore par le signal vocal”

Résumé : La synthèse numérique des sons permet aujourd'hui de reproduire des percepts sonores complexes et réalistes. Les enjeux scientifiques générés par de telles avancées se situent désormais au niveau du contrôle intuitif (ou haut-niveau) de ces percepts. Dans ce contexte, la vocalisation est un moyen puissant pour mettre en relief des caractéristiques sonores complexes, et en particulier les évolutions dynamiques de phénomènes multidimensionnels. En outre, le processus même d'imitation vocale est un formidable moyen d'interroger la perception auditive et de comprendre quels aspects du signal sonore sont porteurs de sens. Ce projet de thèse a pour objectif d'exploiter la puissance et l'intuitivité de la vocalisation en tant que principe de communication ou bien en tant que moyen de contrôle de la synthèse sonore dans le cadre d'une interaction homme-système. Les principaux objectifs de la thèse sont donc : (a) comprendre comment l'imitation vocale permet de révéler des “invariants morphologiques sonores” porteurs de sens ; (b) mettre au point des méthodes d'analyse adaptées au signal vocal afin de définir des descripteurs acoustiques relatifs à ces invariants ; (c) proposer un nouveau paradigme d'interface pour un contrôle intuitif de la synthèse sonore.

Etienne THORET (Octobre 2011 – Décembre 2014) : Financement MENRT

Co-direction avec S. Ystad (LMA)

Titre : “Caractérisation acoustique des relations entre les mouvements biologiques et la perception sonore : application au contrôle de la synthèse et à l'apprentissage de gestes”

Thèse soutenue au LMA le 19 Décembre 2014

Situation actuelle : Post-doctorat à l'Université de McGill (Montréal, Canada)

Résumé : Cette thèse s'est intéressée aux relations entre les mouvements biologiques et la perception sonore en considérant le cas spécifique des mouvements graphiques et des sons de frottement qu'ils génèrent. L'originalité de ces travaux réside dans l'utilisation d'un modèle de synthèse sonore basé sur un principe perceptif issu de l'approche écologique de la perception et contrôlé par des modèles de gestes. Des stimuli sonores dont le timbre n'est modulé que par des variations de vitesse produites par un geste ont ainsi pu être générés permettant de se focaliser sur l'influence perceptive de cet invariant transformationnel. Une première étude a ainsi montré que l'on reconnait la cinématique des mouvements biologiques (la loi en puissance 1/3), et que l'on peut discriminer des formes géométriques simples juste à partir des sons de frottement produits. Une seconde étude a montré l'existence de prototypes dynamiques sonores caractérisant les trajectoires elliptiques, mettant ainsi en évidence que les prototypes géométriques peuvent émerger d'un couplage sensorimoteur. Enfin, une dernière étude a montré qu'une cinématique évoquée par un sonore influence significativement la cinématique et la géométrie

d'un geste dans une tâche de reproduction graphique du mouvement d'un point lumineux. Ce résultat révèle l'importance de la modalité auditive dans l'intégration multisensorielle des mouvements continus dans une situation jamais explorée. Ces résultats ont permis le contrôle de modèles de synthèse par des descriptions gestuelles et la création d'outils de sonification pour l'apprentissage de gestes et la réhabilitation d'une pathologie motrice, la dysgraphie.

Simon CONAN (Novembre 2011 – Décembre 2014) : Financement ANR

Co-direction avec R. Kronland-Martinet (LMA)

Titre : "Contrôle intuitif de la synthèse sonore d'interactions solidiennes – Vers les métaphores sonores"

Thèse soutenue au LMA le 3 Décembre 2014

Situation actuelle : Ingénieur R&D audio à SERATO (Auckland, Nouvelle-Zélande)

Résumé : Un des enjeux actuels de la synthèse sonore est le contrôle perceptif (i.e. à partir d'évocations) des processus de synthèse. En effet, les modèles de synthèse sonore dépendent généralement d'un grand nombre de paramètres de bas niveau dont la manipulation nécessite une expertise des processus génératifs. Disposer de contrôles perceptifs sur un synthétiseur offre cependant beaucoup d'avantages en permettant de générer les sons à partir d'une description du ressenti et en offrant à des utilisateurs non-experts la possibilité de créer et de contrôler des sons intuitivement. Un tel contrôle n'est pas immédiat et se base sur des hypothèses fortes liées à notre perception, notamment la présence de morphologies acoustiques, dénommées "invariants", responsables de l'identification d'un évènement sonore. Cette thèse aborde cette problématique en se focalisant sur les invariants liés à l'action responsable de la génération des sons. Elle s'articule suivant deux parties. La première a pour but d'identifier des invariants responsables de la reconnaissance de certaines interactions continues : le frottement, le grattement et le roulement. Le but est de mettre en œuvre un modèle de synthèse temps-réel contrôlable intuitivement et permettant d'effectuer des transitions perceptives continues entre ces différents types d'interactions (e.g. transformer progressivement un son de frottement en un son de roulement). Ce modèle s'inscrira dans le cadre du paradigme "action-objet" qui stipule que chaque son résulte d'une action (e.g. gratter) sur un objet (e.g. une plaque en bois). Ce paradigme s'adapte naturellement à une approche de la synthèse par modèle source-filtre, où l'information sur l'objet est contenue dans le "filtre", et l'information sur l'action dans la "source". Pour ce faire, diverses approches sont abordées : études de modèles physiques, approches phénoménologiques et tests perceptifs sur des sons enregistrés et synthétisés. La seconde partie de la thèse concerne le concept de "métaphores sonores" en élargissant la notion d'objet à des textures sonores variées. La question posée est la suivante : étant donnée une texture sonore quelconque, est-il possible de modifier ses propriétés intrinsèques pour qu'elle évoque une interaction particulière comme un frottement ou un roulement par exemple ? Pour créer

ces métaphores, un processus de synthèse croisée est utilisé dans lequel la partie “source” est basée sur les morphologies sonores des actions précédemment identifiées et la partie “filtre” restitue les propriétés de la texture. L’ensemble de ces travaux ainsi que le paradigme choisi offre dès lors de nouvelles perspectives pour la constitution d’un véritable langage des sons.

Camille-Eva RAKOVEC (Octobre 2012 – Janv. 2014) : Financement MENRT

Co-direction avec R. Kronland-Martinet (LMA)

Titre : “Perception des objets du quotidien en vue du contrôle intuitif d’un synthétiseur de sons d’impact”

Thèse interrompue en Janvier 2014 pour cause médicale

Résumé : Dans le cadre de la réalité virtuelle, un des enjeux scientifiques actuel réside dans la prise en compte des actions de l’utilisateur dans le processus de synthèse sonore afin d’augmenter la sensation d’immersion et de réalisme. Dans ce contexte, le travail de thèse envisagé a pour objectif d’aborder les problématiques spécifiques aux sons produits par la manipulation d’objets de l’environnement. Il s’agira tout d’abord de mieux comprendre la perception auditive des caractéristiques physiques de ces sources (forme, taille, présence de cavité, etc.) et de définir des catégories perceptives propres à ces caractéristiques par des tests d’écoute. La détermination d’ “invariants” responsables des évocations induites s’appuiera sur des considérations physiques et sur la caractérisation acoustique des signaux produits. Basés sur ces résultats perceptifs et acoustiques, il s’agira par la suite de mettre en œuvre des stratégies de contrôle perceptif et interactif, en vue de piloter un synthétiseur générique de sons d’environnement.

Charles VERRON (Octobre 2006 - Octobre 2009) : Financement CIFRE

Co-direction avec R. Kronland-Martinet (LMA) et G. Pallone (tuteur à Orange Labs)

Titre : “Synthèse immersive de sons d’environnement”

Thèse soutenue au LMA le 7 Mai 2010

Situation actuelle : Ingénieur et co-fondateur de *Noise Makers* (Rennes, France)

Résumé : Ce travail de thèse propose une approche originale pour la synthèse et la spatialisation des sons d’environnement. Nous concevons un synthétiseur temps-réel qui permet de générer des sources sonores diverses (pluie, vent, feu, vagues, impacts, etc.), de les contrôler via des paramètres “haut-niveau”, et de les manipuler dans un espace virtuel 3D. Après une revue des modèles de synthèse existants pour diverses sources environne- mentales, nous spécifions un modèle de synthèse générique basé sur la synthèse additive et sur cinq structures sonores paramétriques appelées “atomes”. Une combinaison adéquate de ces atomes permet de créer des sources associées aux trois catégories de sons d’environnement : les solides en vibration, les sources aérodynamiques et les liquides. Le moteur du synthétiseur intègre efficacement synthèse et spatialisation au même niveau de la génération du son. Nous utilisons l’implémentation

de la synthèse additive par transformée de Fourier inverse avec un choix de paramètres adapté. Le rendu sonore 3D est compatible avec les formats audio standard (multi-canal, ambisonics, binaural) permettant une écoute sur haut-parleurs ou au casque. Nous présentons également une méthode de synthèse basée sur un formalisme en "sous-bande" qui permet de générer des signaux stochastiques nécessitant une résolution fine à la fois en temps et en fréquence (comme des sons d'impact ou de feu). Enfin, des contrôles "haut-niveau" sont spécifiés pour manipuler intuitivement le synthétiseur à partir de propriétés physiques telles que la force du vent ou l'intensité du feu. Nous proposons également un contrôle d'extension spatiale, validé par un test perceptif formel, qui permet de simuler des sources naturellement non ponctuelles comme la pluie ou les vagues. Le contrôle conjoint du timbre et des propriétés spatiales des sources permet de créer intuitivement des scènes sonores complexes et immersives.

Encadrements de stage

- Avril 2016 - Septembre 2016 : Antoine BOURACHOT. Stage de Master 2 Recherche - "Mécanique, Physique et Ingénierie" Mention Acoustique sur le thème "Nouvelles stratégies de contrôle intuitif pour la synthèse sonore", en collaboration avec S. Ystad, R. Kronland-Martinet et L. Pruvost (LMA).
- Mars 2014 - Juillet 2014 : Tristan BEGUERIA. Stage de 2ème année d'Ecole Nationale Supérieure des Mines de Saint-Etienne (ISMIN, Site Georges Charpak, Gardanne) sur le thème "Perception de la forme des objets du quotidien : approche modale en vue du contrôle sémantique de la synthèse sonore", en collaboration avec S. Ystad et R. Kronland-Martinet (LMA).
- Mai 2013 - Août 2013 : Arthur MINGASSON. Stage de 2ème année d'Ecole Centrale Lille sur le thème "Synthèse sonore pour les jeux", en collaboration avec C. Verron (LMA).
- Mars 2012 - Juillet 2012 : Camille-Eva RAKOVEC. Stage de Master 2 Mécanique, Physique et Ingénierie (Spécialité Acoustique) sur le thème "Etude perceptive des objets du quotidien en vue du contrôle intuitif d'un synthétiseur de sons d'impact" en collaboration avec R. Kronland-Martinet et S. Ystad au LMA.
- Mars 2011 - Juillet 2011 : Julien PERON. Stage de fin d'études de l'ENSICAEN (Ecole Nationale Supérieure d'Ingénieurs de Caen) sur le thème "Méthodes haute résolution pour le traitement de signaux audio - Modèle prédictif pour la catégorisation perceptive de matériaux impactés" en collaboration avec R. Kronland-Martinet au LMA.
- Mars 2011 - Juillet 2011 : Etienne THORET. Stage de Master 2 Mécanique, Physique et Ingénierie (Spécialité Acoustique) sur le thème "Vers la sonification des formes : étude de la perception des gestes humains à travers la synthèse sonore de bruits de frottements" en collaboration avec R. Kronland-Martinet et S. Ystad au LMA.

- Mai 2010 - Juillet 2010 : Simon CONAN. Stage de 2ème année Ecole Centrale Marseille sur le thème “Synthèse et contrôle haut-niveau des bruits de pas, application à la réalité virtuelle sonore” en collaboration avec R. Kronland-Martinet et Ch. Verron au LMA.
- Mars 2008 - Août 2008 : Patrick MARMAROLI. Stage de Master 2“Signal et Trajectographie” et de fin d’études de l’ISEN (Université du Sud, Toulon) sur le thème “Catégorisation des sons de matériaux frappés : approches perceptive et statistique” en collaboration avec R. Kronland-Martinet au LMA.
- Mars 2007 - Juin 2007 : Adrien MERER. Stage de Master 2 Mécanique, Physique et Modélisation (Spécialité Acoustique) sur le thème “Vers une approche acoustique et cognitive de la sémiotique des objets sonores. Recherche des invariants morphologiques des sons évoquant un mouvement-déplacement : application à la synthèse” en collaboration avec S. Ystad et R. Kronland-Martinet au LMA.
- Mars 2006 - Juin 2006 : Thibaud NECCIARI. Stage de Master 2 SMES/E2S-Recherche (Université du Sud, Toulon) sur le thème “Analyse-Synthèse de sons impulsifs” en collaboration avec R. Kronland-Martinet et Th. Voinier au LMA.
- Janv. 2006 - Juin 2006 : Céline MARIE. Stage de Master 2 Neurosciences (Spécialité Neurosciences intégratives et cognitives) sur le thème “Approches comportementale et électrophysiologique de la sémiotique des sons”, en collaboration avec M. Besson à l’INCM.

Organisation de manifestations scientifiques

Le groupe dans lequel je travaille est l’instigateur du cycle de conférences CMMR – Computer Music Multidisciplinary Research – dont la première édition a eu lieu à Montpellier en 2003. Depuis, le congrès s’est déplacé à travers l’Europe et même l’Asie, transitant de l’Italie à l’Angleterre en passant par le Danemark, l’Espagne et l’Inde. Au fil des années, CMMR est devenu un événement scientifique international rassemblant des chercheurs de diverses communautés autour des problématiques liées au son, depuis sa production acoustique jusqu’à sa perception et son interprétation, sans oublier les enjeux technologiques tels que la compression ou la navigation dans des bases de données sonores. Je participe activement à ces conférences CMMR depuis 2007 comme membre du comité d’organisation, en particulier comme :

- Proceeding et Paper chair du 13ième *International Symposim on Computer Music Multidisciplinary Research CMMR “Music Technology with Swing”*, 25-28 Septembre 2017, Porto, Portugal. http://cmmr2017.inesctec.pt/
- Proceeding et Paper chair du 12ième *International Symposim on Computer Music Multidisciplinary Research CMMR “Bridging People and Sound”*, 5-8 Juillet 2016, São Paolo, Brésil. http://cmmr2016.ime.usp.br/

- Paper Chair du 11ième CMMR "Music, Mind and Embodiment", 16-19 Juin 2015, ICCMR, Plymouth, UK. http://cmr.soc.plymouth.ac.uk/cmmr2015/
- Paper Chair du 10ième *CMMR "Sound, Music and Motion"*, 15-18 Octobre 2013, LMA, Marseille, France. http://www.cmmr2013.cnrs-mrs.fr/
- Proceeding et Paper chair du 9ième *International Symposium on Computer Music Modeling and Retrieval CMMR "Music & Emotion"*, 19-22 Juin 2012, Queen Mary University of London, Londres, UK. http://www.cmmr2012.eecs.qmul.ac.uk/
- Symposium co-chair et membre du Program Committee du congrès *Frontiers of Research on Speech and Music joint with Computer Music Modeling and Retrieval FRSM/ CMMR 2011 - Singing synthesis*, 9-12 Mars 2011, Université d'Utkal, Bhubaneswar, Inde.
- Paper et Program chair du *CMMR 2010 - Music is in the Sound*, 21-24 Juin 2010, Málaga, Espagne. http://www.cmmr2010.etsit.uma.es/
- Paper chair du *15th International Conference on Auditory Display joint with the 6th International Symposium on Computer Music Modeling and Retrieval CMMR/ICAD 2009*, 18-21 Mai 2009, Copenhague, Danemark.
  http://www.lma.cnrs-mrs.fr/cmmr2009/
- Membre du comité scientifique du *CMMR 2008 - Genesis of Meaning in Sound and Music*, 19-23 Mai 2008, Copenhague, Danemark.
- Membre du comité scientifique du *CMMR 2007 - Sense of Sounds*, 27-31 Août 2007, Copenhague, Danemark. http://www.lma.cnrs-mrs.fr/cmmr2007/

Par ailleurs, j'ai également participé à l'organisation des 2èmes Journées Perception Sonore (JPS), en collaboration avec le Groupe Perception Sonore de la Société Française d'Acoustique, l'Ircam et le GDR Visible en 2012 (10 et 11 décembre, LMA, Marseille ; http://www.lma.cnrs-mrs.fr/jps2012/).

Activités éditoriales

Le cycle de conférences CMMR a donné lieu depuis 2003 à la publication annuelle d'un livre de la série "Lecture Notes in Computer Science Series" (édition Springer Verlag Heidelberg) rassemblant les meilleurs articles scientifiques sélectionnés à l'issue de chaque congrès. Dans ce cadre, j'ai participé à la co-édition des ouvrages suivants :

- *Music, Mind, and Embodiment*. R. Kronland-Martinet, M. Aramaki and S. Ystad, Ed., volume 9617 de Lecture Notes in Computer Science. ISBN : 978-3-319-46281-3 (Print) 978-3-319-46282-0 (Online). Springer Berlin Heidelberg, 2016.

- *Sound, Music, and Motion*. M. Aramaki, O. Derrien, R. Kronland-Martinet and S. Ystad, Ed., volume 8905 de Lecture Notes in Computer Science. ISBN : 978-3-319-12975-4 (Print) 978-3-319-12976-1 (Online). Springer Berlin Heidelberg, 2014.
- *From Sounds to Music and Emotions*. M. Aramaki, M. Barthet, R. Kronland-Martinet and S. Ystad, Ed., volume 7900 de Lecture Notes in Computer Science. ISBN : 978-3-642-41247-9 (Print) 978-3-642-41248-6 (Online). Springer Berlin Heidelberg, 2013.
- *Speech, Sound and Music Processing : Embracing Research in India*. S. Ystad, M. Aramaki, R. Kronland-Martinet, K. Jensen, et S. Mohanty, Ed., volume 7172 de Lecture Notes in Computer Science. ISBN : 978-3-642-31979-2 (Print) 978-3-642-31980-8 (Online). Springer Berlin Heidelberg, 2012.
- *Exploring Music Contents*. S. Ystad, M. Aramaki, R. Kronland-Martinet, et K. Jensen, Ed., volume 6684 de Lecture Notes in Computer Science. ISBN : 978-3-642-23125-4 (Print) 978-3-642-23126-1 (Online). Springer Berlin Heidelberg, 2011.
- *Auditory Display*. S. Ystad, M. Aramaki, R. Kronland-Martinet, et K. Jensen, Ed., volume 5954 de Lecture Notes in Computer Science. ISBN : 978-3-642-12438-9 (Print) 978-3-642-12439-6 (Online). Springer Berlin Heidelberg, 2010.

Transfert technologique, relations industrielles et valorisation

Les projets ANR auxquels j'ai participé ont permis de nombreuses collaborations avec des partenaires académiques (LNC, ISM, IRCAM, LATP) et industriels (Orange Labs, AudioGaming, Game Audio Factory, PCA). En particulier, j'ai eu la responsabilité scientifique pour le LMA dans le projet **Physis** et d'assurer les relations avec les partenaires industriels. Les projets sont listés ci-dessous :

- 2014 - 2018 : Participation au projet ANR **SoniMove** "Informer, guider et influencer l'action par le son" (ANR-14-CE24-0018) coordonné par R. Kronland-Martinet (PRISM, Marseille). Autres partenaires : Institut des Sciences du Mouvement (ISM, Marseille), Peugeot-Citroën Automobiles (PCA, Vélizy-Villacoublay).
  http://sonimove.lma.cnrs-mrs.fr/
- 2014 - 2017 : Participation au projet ANR Blanc Transnational Collaboration Bilateral Agreement (Franco-autrichien) **POTION** "Perceptual optimization of time-frequency audio representations and coding" coordonné par T. Necciari (ARI, Autriche) et O. Derrien (PRISM, Marseille). http://potion.cnrs-mrs.fr/
- 2012 - 2015 : Responsable scientifique pour le LMA dans le projet ANR Programme CONTINT **Physis** "Modèles de synthèse audio interactifs, physiquement informés et sémantiquement

contrôlables" (ANR-12-CORD-0006) coordonné par D. Henry (AudioGaming). Autres partenaires : AudioGaming, Game Audio Factory, IRCAM.
http://physis-project.net

- 2010 - 2014 : Participation au projet ANR Programme CONTINT **MetaSon** "Métaphores Sonores" (ANR-10-CORD-0003) coordonné par R. Kronland-Martinet (LMA, Marseille). Autres partenaires : Laboratoire de Neurosciences Cognitives (LNC, Marseille), Laboratoire d'Analyse, Topologie, et Probabilités (LATP, Marseille), Peugeot-Citroën Automobiles (PCA, Vélizy-Villacoublay). http://metason.cnrs-mrs.fr/
- 2006 - 2009 : Responsable scientifique d'un contrat particulier de recherche externalisée (N° 46142731) entre le LMA, l'INCM, l'Université de la Méditerranée et France-Télécom accompagné d'une bourse CIFRE (N° 168/2007).
- 2006 - 2009 : Participation au projet ANR Jeunes Chercheuses et Jeunes Chercheurs **senSons** "Vers le sens des sons" (ANR-05-JCJC-0033-01) coordonné par S. Ystad (LMA, Marseille). http://sensons.cnrs-mrs.fr/

Une collaboration avec des chercheurs du Centre de coopération Internationale en Recherche Agronomique pour le Développement (CIRAD, Montpellier) ont également permis de mener plusieurs études qui ont été valorisées par des publications.

Distinctions

- 2013 : Prix du meilleur article du colloque international DAFx'13 pour : "Conan S., Thoret E., Aramaki M., Derrien O., Gondre C., Kronland-Martinet R., and Ystad S. Navigating in a Space of Synthesized Interaction-Sounds : Rubbing, Scratching and Rolling Sounds"
- 2012 : Sélection par ACM comme "top best items published in computing in 2012" du livre : "Ystad, S., Aramaki, M., Kronland-Martinet, R. ; Jensen, K. ; and Mohanty, S. (Eds.) Speech, sound and music processing : embracing research in India, LNCS 7172, Springer Verlag Heidelberg."
- 2013, distinction collective : Nomination du projet ANR *Physis* comme projet phare.
- 2012, distinction collective : Nomination du projet ANR *MetaSon* comme projet phare.

Responsabilités collectives et management de la recherche

- De janvier 2012 à Avril 2013, j'ai participé à la vie collective du LMA comme membre élue du Conseil de Laboratoire.

- Le projet ANR *Physis* pour lequel j'ai été responsable scientifique pour le LMA, m'a permis d'animer une équipe de recherches et de gérer le budget alloué au LMA ainsi que les embauches de contractuels (ingénieurs, post-doctorants).
- Je suis membre fondateur de la future UMR PRISM (Perception, Représentations, Image, Son, Musique) qui sera effective au 1er Janvier 2018 et constituée en FRE depuis le 1er Janvier 2017. Au sein de cette structure, je suis responsable de l'axe "Ingénierie de la perception" ainsi que du thème transverse "Perception virtuelle et augmentée - applications industrielles, sociétales et thérapeutiques". Je fais également partie du comité de direction et du comité scientifique de PRISM.

Mobilité

Depuis mon recrutement au CNRS, j'ai effectué un changement de laboratoire en janvier 2012 de l'INCM vers le LMA. La mobilité que j'ai effectuée au cours de mon parcours scientifique a été également thématique. En effet, ma formation initiale m'a permis d'aborder le phénomène sonore du point de vue de la physique et de l'acoustique. Mes travaux de thèse ont porté sur la construction de modèles d'analyse-synthèse de sons d'impact. L'évolution vers les aspects perceptifs et cognitifs s'est produite de façon naturelle lorsque j'ai abordé les problématiques liées à la validation et la calibration de ces modèles de synthèse dans le contexte de la réalité virtuelle et augmentée. Cette évolution s'est traduite par mon intégration dans un laboratoire de neurosciences cognitives, à l'INCM, qui m'a permis de développer un projet de recherches à l'interface entre l'acoustique et les neurosciences cognitives. La mise en place d'un tel projet a impliqué des prises de risque inhérentes à l'interdisciplinarité tout en préservant un lien étroit avec ma discipline d'origine. Parmi ces risques, on peut noter une période incompressible d'adaptation à une nouvelle culture scientifique, une perte de visibilité au sein de la discipline d'origine et une difficulté à faire expertiser et à publier des travaux fondamentalement pluridisciplinaires.

Sociétés savantes

- Membre de la Société Française d'Acoustique (SFA)
- Membre de Audio Engineering Society (AES)
- Membre senior de Institute of Electrical and Electronics Engineers (IEEE)

Cinquième partie

Projet de recherche

“Un langage des sons pour une ingénierie de la perception”

Les travaux que j’ai menés jusqu’à présent constituent les fondations d’un véritable langage des sons dont le contenu informatif serait contrôlable par la synthèse. Une des perspectives majeures de ces travaux concernera la formalisation de ce langage et du paradigme d’étude adopté ainsi que le développement de nouvelles méthodologies permettant d’évaluer l’apport de ce nouveau langage dans différents contextes.

Mes recherches seront menées dans la future UMR PRISM (Perception, Représentations, Image, Son, Musique) qui sera créée au 1er Janvier 2018 dans le cadre de la vague C de l’H-CERES, mais déjà constituée en Formation de Recherche en Evolution (FRE) depuis le 1er Janvier 2017. J’ai activement œuvré à la création de ce laboratoire qui offre un cadre interdisciplinaire hors du commun pour le développement de mes perspectives de recherches. Il a pour objectif de rassembler des forces de recherche existantes et collaborant depuis de nombreuses années autour de problématiques liées à l’ingénierie de la perception (ingénierie de la santé, réalité virtuelle et augmentée, langage des sons, etc.) et aux représentations (analyse, procédures, dispositifs, interactions, etc.). L’enjeu du laboratoire repose sur une forte interdisciplinarité axée sur des objets de recherche partagés : le son, l’image et la musique. Il établit une transversalité entre Sciences, Technologies et Arts et débouche sur des actions et des résultats dans les domaines de la recherche fondamentale et appliquée, en liaison étroite avec le monde professionnel.

Trois axes de recherche constituent les fondations du laboratoire, traitants de questions fondamentales au cœur de débats scientifiques, technologiques et artistiques actuels. Au sein de cette structure, j’assure la responsabilité de l’axe “Ingénierie de la perception” qui adresse les problématiques liées à la perception au travers des nouvelles technologies. Cet axe compte une dizaine de permanents (chercheurs et enseignant-chercheurs) et une dizaine de non permanents (doctorants, post-doc, CDD). Il aborde à la fois les questions fondamentales relatives aux interactions multimodales, en particulier les interactions entre l’audition, la vision et le mouvement, mais aussi les questions pragmatiques liées à l’apport des nouvelles technologies pour l’augmentation des perceptions. Cet axe repose sur une dualité a priori antinomique mais qui correspond

bien à une action transdisciplinaire émergente : *l'ingénierie de la perception*. Les retombées des recherches effectuées dans le cadre de cet axe de recherche seront nombreuses et touchent aux questions sociétales et industrielles liées aux nouvelles technologies embarquées et aux systèmes immersifs et augmentés. Cet axe contiendra également, par le biais de l'ingénierie de la santé et les vécus de conscience, une ouverture sur le monde médical et les problématiques de thérapeutiques non invasives et non chimiques. Cet axe sera déployé selon les thématiques de recherche transversaux décrites ci-dessous.

A Le langage des sons

Ce thème vise à la construction d'un véritable langage des sons. Il s'agit ici de formaliser le contexte théorique du paradigme d'étude développé jusqu'à présent. Un tel objectif implique de repenser la représentation de notre monde sonore en croisant les connaissances mathématiques, physiques et vibratoires aux connaissances perceptives et cognitives. Ce thème reprend les nombreuses questions fondamentales abordées dans les travaux de recherche effectués, liées notamment au sens attribué aux sons, aux morphologies d'un signal sonore responsable de l'attribution du sens, aux représentations mathématiques et perceptives des sons, à l'influence des sons sur le comportement humain et au rôle de la perception auditive en présence d'autres modalités sensorielles. Il s'agira par ailleurs de développer et de pérenniser la plateforme de synthèse permettant de créer et transformer des sons réalistes ou inouïs et de contrôler les sons de façon continue et intuitive. Les applications de ce travail seront nombreuses : domaine de l'automobile (sonification de voitures électriques et de nouveaux usages), de l'aéronautique (alarmes dans les cockpits d'avion), de la création musicale (nouveaux paradigmes d'exploration de l'espace sonore, sons inouïs) et du cinéma (doublage son automatique pour la voix et les sons d'environnement, indexation supervisée, clonage de locuteurs).

A.1 Représentation et modélisation des sons

La caractérisation et la manipulation des signaux sonores sont fortement dépendantes de la forme mathématique utilisée pour représenter le signal. Les travaux viseront à proposer de nouvelles alternatives aux méthodes classiques en intégrant des concepts perceptifs. Nos efforts vont consister à définir de nouvelles représentations ou à perfectionner des représentations existantes en s'appuyant sur les nouvelles méthodes de décomposition adaptative avec contraintes perceptive et de parcimonie, ainsi que les nouvelles transformées non-stationnaires perceptuelles, en collaboration avec l'Institut de Mathématiques de Marseille (B. Torrésani). Nous continuerons également à développer les décompositions du signal en modes oscillants et amortis, rendues possibles par les nouvelles méthodes haute-résolution (dérivées de l'algorithme ESPRIT), ainsi que les identifications de systèmes source-filtre complexes (utilisant des modèles statistiques

évolués de l'excitation). Ces modèles permettront d'identifier les paramètres physiques et acoustiques des modèles de synthèse. Enfin, il s'agira d'approfondir les recherches sur les invariants morphologiques afin d'en identifier de nouveaux, en explorant différentes approches d'investigation de la perception humaine telles que l'imitation vocale et gestuelle ainsi que les entretiens d'explicitation.

A.2 Synthèse et contrôle perceptif des sons

Sur la base du paradigme {action-objet} décrit dans les travaux de recherche effectués, il s'agira ici de construire des modèles de synthèse sonore permettant la manipulation d'invariants morphologiques sonores. L'objectif final est le contrôle des sons à haut niveau, basé sur une organisation multidimensionnelle des sons portée par des descriptions sémantiques reliés aux évocations induites par les sons, en accord avec nos propres modèles cognitifs. Le paradigme de contrôle des sons permet l'hybridation d'invariants "improbables" et la construction d'objets sonores inouïs (frotter le vent, faire couler du métal, etc.). Afin d'investiguer les potentialités d'une telle approche à la construction d'un véritable langage des sons, la plateforme de synthèse interactive et temps réel continuera d'être développée. Elle intègrera les méthodologies qui seront mises en place et constituera le cœur pérenne de nos réalisations. Elle sera également à la base des études sur les interactions multimodales et sur les environnements virtuels.

Nous aborderons spécifiquement la problématique de l'exploration des vécus de conscience dans l'écoute des sons inouïs et dans les processus de contrôle intuitif. En effet, se pose la question de la manière dont peut s'effectuer le contrôle de la production sonore non pas en considérant a priori les paramètres physiques du son, mais en passant par la description de la boucle perceptive dans laquelle les processus conscients sont impliqués principalement de manière non réflexive. L'analyse des vécus de conscience s'avère très informative quant aux représentations mentales générées par la perception du son et quant aux processus d'incarnation, i.e. mouvements potentiels suscités par l'écoute.

A.3 Aspect multimodal de la perception

Comme pour tout langage, la "signification" des sons dépend fortement du contexte. L'étude de l'influence de l'environnement sonore sur la perception sonore et le comportement humain et de l'interaction avec les autres modalités, représente un fort enjeu des recherches envisagées ici. En exploitant les attributs propres à l'information auditive, le son constitue un moyen privilégié d'influence sur l'humain. Nous avons pu montrer qu'une cinématique évoquée par un son influence significativement la cinématique du geste et la trace produite dans une tâche de reproduction graphique d'un mouvement visuel. Ce résultat révèle l'importance de la modalité auditive dans l'intégration multisensorielle des mouvements continus. Il s'agira ainsi d'évaluer

la manière dont le contrôle de la morphologie de sons permet d'agir de façon explicite ou implicite sur notre comportement et d'établir des modèles empiriques et/ou formels d'interactions multimodales mettant en jeu la modalité auditive. Nous poursuivrons les investigations entreprises dans le projet SoniMove sur les nouvelles méthodes d'apprentissage par le son allant du domaine sportif au domaine thérapeutique ou éducatif, en collaboration avec les enseignants-chercheurs de l'ISM (C. Bourdin et L. Bringoux). Ainsi, nous étudierons les liens naturels entre gestes et sons afin de définir des stratégies génériques de guidage de gestes. Nous viserons à formaliser des modèles d'interactions afin de calibrer les manipulations sonores aptes à orienter le geste vers un objectif spécifique.

B Perception virtuelle et augmentée

Ce thème adresse les problématiques liées aux nouvelles formes de perception au travers des technologies émergentes, qu'elles soient embarquées ou associées à des dispositifs immersifs et interactifs. Ces technologies fournissent une grande variété de données physiques, qui en les rendant sensibles et signifiantes, permettraient d'accéder à des perceptions inédites du monde. Ces aspects soulèvent des questions fondamentales communes aux applications envisagées : Quelles informations doit-on révéler et comment les révéler, notamment par le son ? Comment notre perception est-elle modifiée par l'adjonction d'informations sensorielles multiples via des technologies ? Ils appellent également à une réflexion nouvelle et générale sur l'interaction entre l'homme, la science et la technologie, et en particulier sur l'appropriation de ces nouveaux objets susceptibles d'être appréhendés comme des "greffes" technologiques.

Ce thème vise ainsi à anticiper le bouleversement culturel et environnemental induit par une augmentation de nos perceptions, notamment sous-tendue par le son, selon différents cas d'usages dans le domaine industriel (transport, sport, etc.), sociétal (apprentissage, musique, cinéma, etc.) et de l'ingénierie pour la santé (remédiation, biofeedback). Les recherches s'appuieront sur la mise en place d'une plateforme technologique d'immersion multisensorielle visant à étudier le son dans un contexte multimodal large (audition, vision, proprioception).

B.1 Technologies émergentes embarquées

Les nouvelles interfaces de commande par tablette ou par le geste offrent de nombreuses perspectives dans l'interaction avec des systèmes de plus en plus complexes. Dans certaines situations (de conduite par exemple), elles peuvent être coûteuses en ressources cognitives attentionnelles, en particulier visuelles. Les travaux menés actuellement dans le projet ANR SoniMove concernent le développement de nouvelles interfaces homme-machine sonores, l'aide à l'apprentissage d'un geste expert précis, à savoir le jeu instrumental dans le contexte musical

et le geste sportif. Les innovations proposées visent à une véritable évolution des pédagogies d'apprentissage qui seront évaluées dans ces domaines. Nous nous intéresserons au développement d'équipements sonifiés pour apprendre un geste expert précis. Enfin, dans le domaine clinique, les aides auditives s'équipent de technologies miniaturisées intégrant des procédés de traitement du signal évolués mais pas toujours basés sur la prise en compte de critères perceptifs. Il s'agira ici d'explorer de nouvelles pistes d'encodage de l'information utile basées sur les invariants morphologiques sonores relatives à l'évocation des sources. En effet, un nouveau codage de l'implant cochléaire basé sur des informations sémiotiquement robustes pourra être évalué en terme d'efficacité du dispositif et de performance perceptivo-cognitive (intelligibilité, compréhension du sens).

Le développement à la fois d'interfaces hommes-machines, d'équipements sonifiés, de mondes virtuels immersifs (cinéma 3D, d'environnements audio-visuels interactifs) va interroger le chercheur sur les processus perceptifs et les moyens d'appropriation de ces nouvelles technologies par la conscience. En d'autres termes on peut se demander comment le schéma vécu du corps se modifie en présence de ces technologies, et ce en fonction de chaque type d'expérience, et réciproquement si ces technologies acquièrent un statut d'appropriation proche de celui d'un instrument de musique ou un statut spécifique différent à l'intérieur de la conscience. Il s'agira ici de poser, par le biais de l'analyse des vécus de conscience, le problème de l'incarnation de ces technologies innovantes dont on dit qu'elles augmentent l'homme. La description des vécus de conscience pourrait soulever des hypothèses sur la manière dont les technologies peuvent révéler des potentialités inhérentes à la nature humaine.

B.2 Remédiation par le son des déficiences sensorielles

Le son peut également constituer un outil efficace pour faciliter le diagnostic et la rééducation de déficiences motrices. Une première application en collaboration avec les chercheurs du Laboratoire de Neurosciences Cognitives (J.L. Velay et J. Danna) a été mise en œuvre dans le cadre du projet ANR MetaSon. Les études menées ont permis d'apporter une preuve de concept pour l'utilisation du son dans le cadre du diagnostic et de la remédiation de troubles graphomoteurs, notamment dans la dysgraphie, trouble spécifique dans la production de l'écriture. Le projet proposé vise à consolider et formaliser la méthodologie qui a été mise en place, en finalisant les expériences en cours et en démontrant les potentialités applicatives du son comme vecteur d'information.

Par ailleurs, parmi les technologies supplétives, les dispositifs de substitution sensorielle visent à aider les personnes ayant une déficience sensorielle en convertissant les informations d'un organe sensoriel déficient (par ex. la vision) en informations spécifiques à une autre modalité sensorielle (par ex. le toucher ou l'audition). D'un point de vue fondamental, ces dispositifs permettent d'étudier les mécanismes perceptifs se constituant grâce à une nouvelle forme d'in-

teraction avec l'environnement, aux niveaux de l'acquisition de nouvelles habiletés perceptives, de la phénoménologie associée et de la plasticité structurelle et fonctionnelle du système nerveux central. Les travaux envisagés dans ce cadre sont en cours de discussion avec M. Auvray (ISIR, Paris). Ils se focaliseront sur la substitution visuo-auditive et s'inscriront à la fois en amont et en aval de la conception des dispositifs. Ils impliqueront l'étude de la nature et des possibilités offertes par la substitution sensorielle, des facteurs centraux à leur apprentissage, et leur adéquation à la population cible en prenant en compte les spécificités individuelles.

B.3 Biofeedback basé sur le son

Les méthodes visant à développer une prise de conscience et un contrôle volontaire d'un paramètre physiologique sont dénommées biofeedback (BF, pour l'ECG, l'électrodermogramme, le pneumogramme) et neurofeedback (NF, pour les activités cérébrales : EEG, potentiels évoqués, activations en IRMf). Elles présentent un développement exponentiel en raison de la capacité grandissante des ordinateurs à réaliser des analyses rapides des signaux électrophysiologiques tout en présentant en même temps des stimuli de plus en plus variés à l'origine de la boucle de feedback. Sur un plan clinique, on observe que les indications validées de ces méthodes, et plus particulièrement le NF, sont d'une part les déficits de l'attention du Trouble avec Déficit Attentionnel et Hyperactivité (TDAH), certaines épilepsies pharmaco-résistantes, et les troubles du sommeil. Les traitements par NF sont caractérisés par une absence d'effets secondaires, par une très bonne tolérance auprès des patients, et par l'augmentation de leur qualité de vie. Cependant, les incontestables améliorations constatées en NF sont variables selon les sujets, pas toujours pérennes et nécessitent souvent des séances d'entretien.

L'étude que nous souhaitons mener consiste à améliorer l'efficacité et la pérennité des effets du NF en reconsidérant les procédures de stimulation sensorielle, en particulier auditive, en utilisant les procédés de synthèse et de contrôle intuitif et interactif des sons possibles avec le synthétiseur. Les travaux seront menés en collaboration avec les chercheurs du SANPSY (USR CNRS 3413, Bordeaux), en particulier avec le Dr. J.A. Micoulaud-Franchi. A terme, il s'agirait de concevoir une thérapeutique ambulatoire embarquée par NF, basée sur une miniaturisation du système de NF sur smartphone. Ceci permettrait l'autonomie des personnes et le maintien du lien social pendant les périodes de cure. Les processus d'apprentissage ou de choix de stratégies du NF seront associés à la description des processus conscients réflexifs ou non réflexifs mis en jeu dans ces adaptations à ces modalités techniques.

Bibliographie

Aramaki, M., Baillères, H., Brancheriau, L., Kronland-Martinet, R., and Ystad, S. (2007a). Sound quality assessment of wood for xylophone bars. *Journal of the Acoustical Society of America*, 121(4) :2407–2420.

Aramaki, M. and Besson, M. (2007). Approche électrophysiologique de la sémiotique des sons. In *Musique & Sciences Cognitives, Actes des troisièmes Rencontres Interartistiques de l'Observatoire Musical Français*, Conférences et Séminaires, N° 46, pages 99–107.

Aramaki, M., Besson, M., Kronland-Martinet, R., and Ystad, S. (2009a). Timbre perception of sounds from impacted materials : behavioral, electrophysiological and acoustic approaches. In Ystad, S., Kronland-Martinet, R., and Jensen, K., editors, *Computer Music Modeling and Retrieval - Genesis of Meaning of Sound and Music*, volume 5493 of *LNCS*, pages 1–17. Springer-Verlag Berlin Heidelberg.

Aramaki, M., Besson, M., Kronland-Martinet, R., and Ystad, S. (2011). Controlling the perceived material in an impact sound synthesizer. *IEEE Transactions on Audio, Speech, and Language Processing*, 19(2) :301–314.

Aramaki, M., Brancheriau, L., Baillères, H., Kronland-Martinet, R., and Ystad, S. (2006a). Relationship between sound classification of xylophone-like bars and wood species properties. In Eberhardsteiner, J., Mang, H., and Waubke, H., editors, *Proceedings of the 13th International Congress on Sound and Vibration (ICSV13)*, Vienne, Autriche. Vienna University of Technology, Austria. 2-6 juillet 2006.

Aramaki, M., Brancheriau, L., Kronland-Martinet, R., and Ystad, S. (2008). Perception of impacted materials : sound retrieval and synthesis control perspectives. In Jensen, K., editor, *Proceedings of the Computer in Music Modeling and Retrieval - Genesis of Meaning in Sound and Music*, pages 1–8, Copenhagen, Denmark. Re :New - Digital Arts Forum.

Aramaki, M., Brancheriau, L., Kronland-Martinet, R., and Ystad, S. (2009b). Perception of impacted materials : sound retrieval and synthesis control perspectives. In Ystad, S., Kronland-Martinet, R., and Jensen, K., editors, *Computer Music Modeling and Retrieval - Genesis of Mea-*

ning of Sound and Music, volume 5493 of *LNCS*, pages 134–146. Springer-Verlag Berlin Heidelberg.

Aramaki, M., Gondre, C., Kronland-Martinet, R., Voinier, T., and Ystad, S. (2009c). Thinking the sounds : an intuitive control of an impact sound synthesizer. In *Proceedings of the 15th International Conference on Auditory Display (ICAD 2009)*, pages 119–124, Copenhague, Danemark. 18-21 May 2009.

Aramaki, M. and Kronland-Martinet, R. (2006). Analysis-synthesis of impact sounds by real-time dynamic filtering. *IEEE Transactions on Audio, Speech, and Language Processing*, 14(2) :695–705.

Aramaki, M., Kronland-Martinet, R., Voinier, T., and Ystad, S. (2006b). A percussive sound synthetizer based on physical and perceptual attributes. *Computer Music Journal*, 30(2) :32–41.

Aramaki, M., Kronland-Martinet, R., Voinier, T., and Ystad, S. (2007b). Timbre control of a real-time percussive synthesizer. In *Proceedings of the 19th International Congress on Acoustics (CD-ROM) ; ISBN : 84-87985-12-2*, Madrid, Spain.

Aramaki, M., Kronland-Martinet, R., and Ystad, S. (2012). Perceptual control of environmental sound synthesis. In Ystad, S., Aramaki, M., Kronland-Martinet, R., Jensen, K., and Mohanty, S., editors, *Speech, Sound and Music Processing : Embracing Research in India*, volume 7172 of *LNCS*, pages 172–186. Springer Berlin Heidelberg.

Aramaki, M., Marie, C., Kronland-Martinet, R., Ystad, S., and Besson, M. (2010). Sound categorization and conceptual priming for nonlinguistic and linguistic sounds. *Journal of Cognitive Neuroscience*, 22(11) :2555–2569.

Ballas, J. (1993). Common factors in the identification of an assortment of brief everyday sounds. *Journal of Experimental Psychology : Human Perception and Performance*, 19(2) :250–267.

Caetano, M. and Rodet, X. (2011). Sound morphing by feature interpolation. In *Proceedings of the International Conferance on Acoustics, Speech and Signal Processing*, pages 161–164.

Chowning, J. (1971). The simulation of moving sound sources. *Journal of Audio Engineering Society*, 19(1) :2–6.

Conan, S. (2014). *Contrôle intuitif de la synthèse sonore d'interactions solidiennes – vers les métaphores sonores –*. PhD thesis, Ecole Centrale de Marseille.

Conan, S., Aramaki, M., Kronland-Martinet, R., Thoret, E., and Ystad, S. (2012). Perceptual differences between sounds produced by different continuous interactions. In *Proceedings of the 11th Congrès Français d'Acoustique*, pages 409–414, Nantes, France.

Conan, S., Aramaki, M., Kronland-Martinet, R., and Ystad, S. (2013a). *Post-proceedings 9th International Symposium on Computer Music Modeling and Retrieval (CMMR 2012)*, volume 7900 of *Lecture Notes in Computer Science*, chapter Intuitive Control of Rolling Sound Synthesis. Springer Berlin Heidelberg.

Conan, S., Derrien, O., Aramaki, M., Ystad, S., and Kronland-Martinet, R. (2014a). A synthesis model with intuitive control capabilities for rolling sounds. *IEEE Transactions on Audio, Speech, and Language Processing*, 22(8) :1260–1273.

Conan, S., Thoret, E., Aramaki, M., Derrien, O., Gondre, C., Kronland-Martinet, R., and Ystad, S. (2013b). Navigating in a space of synthesized interaction-sounds : rubbing, scratching and rolling sounds. In *Proceedings of the 16th International Conference on Digital Audio Effects (DAFx-13)*, Maynooth, Ireland.

Conan, S., Thoret, E., Aramaki, M., Derrien, O., Gondre, C., Kronland-Martinet, R., and Ystad, S. (2014b). An intuitive synthesizer of continuous interaction sounds : Rubbing, scratching and rolling. *Computer Music Journal*, 38(4) :24–37.

Conan, S., Thoret, E., Gondre, C., Aramaki, M., Kronland-Martinet, R., and Ystad, S. (2013c). An intuitive synthesizer of sustained interaction sounds. In *Proceedings of the 10th International Symposium on Computer Music Multidisciplinary Research (CMMR 2013) "Sound, Music and Motion"*, Marseille, France.

Danna, J., Fontaine, M., Gondre, C., Thoret, E., Aramaki, M., Kronland-Martinet, R., Ystad, S., and Velay, J. (2013a). Sonifier le geste d'écriture peut-il faciliter son apprentissage ? In *Actes du 15ème Congrès International de l'ACAPS*, Grenoble, France.

Danna, J., Fontaine, M., Paz-Villagrán, V., Gondre, C., Thoret, E., Aramaki, M., Kronland-Martinet, R., Ystad, S., and Velay, J. (2015a). The effect of real-time auditory feedback on learning new characters. *Human Movement Science*, 43 :216–228.

Danna, J., Paz-Villagrán, V., Gondre, C., Aramaki, M., Kronland-Martinet, R., Ystad, S., and Velay, J. (2015b). "let me hear your handwriting !" evaluating the movement quality from its sonification. *Plos One*, 10(6) :e0128388.

Danna, J., Paz-Villagrán, V., Velay, J., Gondre, C., Kronland-Martinet, R., Ystad, S., Aramaki, M., Thoret, E., Conan, S., Voinier, T., Omer, H., and Torrésani, B. (2012a). Sonifier l'écriture : un

outil pour le diagnostic et la remédiation de la dysgraphie. In *IIIèmes Journée Scientifique du Centre de Référence des Troubles d'Apprentissage (CERTA)/RESODYS*, Marseille, France.

Danna, J., Paz-Villagrán, V., Velay, J., Gondre, C., Kronland-Martinet, R., Ystad, S., Aramaki, M., Thoret, E., Conan, S., Voinier, T., Omer, H., and Torrésani, B. (2012b). Sonifier l'écriture : un outil pour le diagnostic et la remédiation de la dysgraphie. *Développements*, 12 :32–40.

Danna, J., Velay, J., Paz-Villagrán, V., Capel, A., Petroz, C., Gondre, C., Thoret, E., Aramaki, M., Ystad, S., and Kronland-Martinet, R. (2013b). Handwriting movement sonification for the rehabilitation of dysgraphia. In *Proceedings of the 10th International Symposium on Computer Music Multidisciplinary Research (CMMR 2013)*.

Danna, J., Velay, J., Paz-Villagrán, V., Capel, A., Petroz, C., Gondre, C., Thoret, E., Aramaki, M., Ystad, S., and Kronland-Martinet, R. (2014). Movement sonification for the diagnosis and the rehabilitation of graphomotor disorders. In *Sound, Music and Motion, Post-proceedings of 10th International Symposium on Computer Music Modeling and Retrieval (CMMR 2013)*. Springer Berlin Heidelberg.

de Saussure, F. (1916). *Cours de linguistique générale*. Payot.

El-Kaim, A., Aramaki, M., Ystad, S., Kronland-Martinet, R., Cermolacce, M., Naudin, J., Vion-Dury, J., and Micoulaud-Franchi, J. (2015). On the correlation between perceptual inundation caused by realistic immersive environmental auditory scenes and the sensory gating inventory in schizophrenia. *European Psychiatry*, 30(5) :606–614.

Frey, A., Aramaki, M., and Besson, M. (2014). Conceptual priming for realistic auditory scenes and for auditory words. *Brain and Cognition*, 84(1) :141–152.

Gaver, W. W. (1993). What in the world do we hear ? an ecological approach to auditory source perception. *Ecological Psychology*, 5(1) :1–29.

Gibson, J. J. (1986). *The Ecological Approach to Visual Perception*. Lawrence Erlbaum Associates.

Giordano, B. L. and McAdams, S. (2006). Material identification of real impact sounds : Effects of size variation in steel, wood, and plexiglass plates. *Journal of the Acoustical Society of America*, 119(2) :1171–1181.

Gobin, P., Kronland-Martinet, R., Lagesse, G. A., Voinier, T., and Ystad, S. (2003). *From Sounds to Music : Different Approaches to Event Piloted Instruments*, volume 2771 of *Lecture Notes in Computer Science*, pages 225–246. Springer Verlag.

Gounaropoulos, A. and Johnson, C. (2006). Synthesising timbres and timbre-changes from adjectives/adverbs. In *Applications of Evolutionary Computing*, pages 664–675. Springer Berlin Heidelberg.

Grey, J. M. (1977). Multidimensional perceptual scaling of musical timbres. *Journal of the Acoustical Society of America*, 61(5) :1270–1277.

Hermes, D. J. (1998). Synthesis of the sounds produced by rolling balls. Internal IPO report no. 1226, IPO, Center for User- System Interaction, Eindhoven, The Netherlands.

Hetrick, W., Erickson, M., and Smith, D. (2012). Phenomenological dimensions of sensory gating. *Schizophrenia Bulletin*, 38 :178–191.

Hoffman, M. and Cook, P. (2006). Feature-based synthesis : Mapping from acoustic and perceptual features to synthesis parameters. In *Proceedings of the International Computer Music Conference*, New Orleans, USA.

Hoffman, M. and Cook, P. (2007). Real-time feature-based synthesis for live musical performance. In *Proceedings of the 7th international conference on New interfaces for musical expression (NIME'07)*, pages 309–312.

Hommel, B., Müsseler, J., Aschersleben, G., and Prinz, W. (2001). The theory of event coding (tec) : A framework for perception and action planning. *Behavioral and Brain Sciences*, 24 :849–878.

Houben, M. (2002). *The sound of rolling objects, perception of size and speed*. PhD thesis, Technische Universiteteit, Eindhoven.

Jehan, T. and Schoner, B. (2001). An audio-driven, spectral analysis-based, perceptual synthesis engine. In *Audio Engineering Society Convention 110*.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics*, 14(2) :201–211.

Kalénine, S., Cheam, C., Izard, V., and Gentaz, E. (2013). Adults and 5-year-old children draw rectangles and triangles around a prototype but not in the golden ratio. *British Journal of Psychology*, 104(3) :400–412.

Kramer, G. (1994). *Auditory Display - Sonification, Audification, and Auditory Interfaces*. Addison-Wesley.

Kronland-Martinet, R., Guillemain, P., and Ystad, S. (1997). Modelling of natural sounds by time-frequency and wavelet representations. *Organised Sound*, 2(3) :179–191.

Kutas, M. and Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, 12 :161–163.

Lacquaniti, F., Terzuolo, C., and Viviani, P. (1983). The law relating the kinematic and figural aspects of drawing movements. *Acta Psychologica*, 54(1-3) :115–130.

Larsson, P. (2010). Tools for designing emotional auditory driver-vehicle interfaces. In *Auditory Display*, Lecture Notes In Computer Science. Springer Berlin Heidelberg.

Marelli, D., Aramaki, M., Kronland-Martinet, R., and Verron, C. (2010a). Time-frequency synthesis of noisy sounds with narrow spectral components. *IEEE Transactions on Audio, Speech, and Language Processing*, 18(8) :1929–1940.

Marelli, D., Aramaki, M., Kronland-Martinet, R., and Verron, C. (2010b). A unified time-frequency method for synthesizing noisy sounds with short transients and narrow spectral components. In *Proceedings of the 35th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 29–32, Dallas, Etats-Unis. IEEE.

Marelli, D., Aramaki, M., Kronland-Martinet, R., and Verron, C. (2012). An efficient time-frequency method for synthesizing noisy sounds with short transients and narrow spectral components. *IEEE Transactions on Audio, Speech, and Language Processing*, 20(4) :1400–1408.

Mariette, N. (2010). Navigation performance effects of render method and latency in mobile audio augmented reality. In *Auditory Display*. Springer Berlin Heidelberg.

Mathews, M. V. (1963). The digital computer as a musical instrument. *Science*, 142(3592) :553–557.

McAdams, S. (1993). *Recognition of sound sources and events*, pages 146–198. Oxford University Press.

McGhie, A. and Chapman, J. (1961). Disorders of attention and perception in early schizophrenia. *British Journal of Medical Psychology*, 34 :103–116.

Merer, A., Aramaki, M., Ystad, S., and Kronland-Martinet, R. (2013). Perceptual characterization of motion evoked by sounds for synthesis control purposes. *Association for Computing Machinery, Transactions on Applied Perception*, 10(1) :1–24.

Merer, A., Ystad, S., Kronland-Martinet, R., and Aramaki, M. (2008). Semiotics of sounds evoking motions : Categorization and acoustic features. In *Computer Music Modeling and Retrieval - Sense of Sounds*, volume 4969 of *Lecture Notes in Computer Science*, pages 139–158. R. Kronland-Martinet and S. Ystad and K. Jensen, Copenhague, Danemark, springer berlin / heidelberg edition.

Merer, A., Ystad, S., Kronland-Martinet, R., and Aramaki, M. (2010). On the potentiality of abstract sounds in perception research. In *Proceedings of the 7th International Symposium on*

Computer Music Modeling and Retrieval (CMMR 2010) - Music is in the sound, pages 207–219, Málaga, Espagne.

Merer, A., Ystad, S., Kronland-Martinet, R., Aramaki, M., Besson, M., and Velay, J. (2007). Perceptual categorization of moving sounds for synthesis application. In *Proceedings of the International Computer Music Conference (ICMC'07)*, Copenhague, Danemark. 27-31 Août 2007.

Micoulaud-Franchi, J., Aramaki, M., Merer, A., Cermolacce, M., Ystad, S., Kronland-Martinet, R., Naudin, J., and Vion-Dury, J. (2012). Towards an exploration of feeling of strangeness in schizophrenia : Perspectives on acousmatic and everyday listening. *Journal of Abnormal Psychology*, 121(3) :628–640.

Micoulaud-Franchi, J., Aramaki, M., Merer, A., Cermolacce, M., Ystad, S., Kronland-Martinet, R., and Vion-Dury, J. (2011). Categorization and timbre perception of environmental sounds in schizophrenia. *Psychiatry Research*, 189(1) :149–152.

Micoulaud-Franchi, J., Aramaki, M., Richieri, R., Cermolacce, M., Faget, C., Ystad, S., Kronland-Martinet, R., Lancon, C., and Vion-Dury, J. (2014a). Effects of clozapine on perceptual abnormalities and sensory gating. a preliminary cross-sectional study in schizophrenia. *Journal of Clinical Psychopharmacology*, 35(2) :184–187.

Micoulaud-Franchi, J., Hetrick, W., Aramaki, M., Bolbecker, A., Boyer, L., Ystad, S., Kronland-Martinet, R., Richieri, R., Faget, C., Faugere, M., El-Kaim, A., Cermolacce, M., Lancon, C., and Vion-Dury, J. (2014b). Do schizophrenia patients with low p50-suppression report more perceptual anomalies with the sensory gating inventory ? *Schizophrenia Research*, 157(1-3) :157–162.

Micoulaud-Franchi, J., Hetrick, W., Boyer, L., Bolbecker, A., Aramaki, M., Ystad, S., Richieri, R., El-Kaim, A., Faget, C., Faugere, M., Cermolacce, M., Kronland-Martinet, R., Lancon, C., and Vion-Dury, J. (2014c). Validation of the french sensory gating inventory : A confirmatory factor analysis. *Psychiatry Research*, 220(3) :1106–1112.

Moog, R. (1987). Position and force sensors and their application to keyboards and related controllers. In New York, A. E. S., editor, *Proceedings of the AES 5th International Conference : Music and Digital Technology*, pages 179–181.

Orgs, G., Lange, K., Dombrowski, J., and Heil, M. (2006). Conceptual priming for environmental sounds and words : An erp study. *Brain and Cognition*, 62(3) :267–272.

Parseihian, G., Gondre, C., Aramaki, M., Kronland-Martinet, R., and Ystad, S. (2013). Exploring the usability of sound strategies for guiding task : toward a generalization of sonification de-

sign. In *Proceedings of the 10th International Symposium on Computer Music Multidisciplinary Research (CMMR 2013)*.

Parseihian, G., Gondre, C., Aramaki, M., Ystad, S., and Kronland-Martinet, R. (2016). Comparison and evaluation of sonification strategies for guidance tasks. *IEEE Transactions on Multimedia*, 18(4) :674–686.

Parseihian, G., Ystad, S., Aramaki, M., and Kronland-Martinet, R. (2015). The process of sonification design for guidance tasks. *Audio Mobility*, 9(2).

Plante, E., van Petten, C., and Senkfor, A. J. (2000). Electrophysiological dissociation between verbal and nonverbal semantic processing in learning disabled adults. *Neuropsychologia*, 38(13) :1669–1684.

Pruvost, L., Scherrer, B., Aramaki, M., Ystad, S., and Kronland-Martinet, R. (2015). Perception-based interactive sound synthesis of morphing solids' interactions. In *Proceedings of the Siggraph Asia 2015*, Kobe, Japon.

Rakovec, C., Aramaki, M., and Kronland-Martinet, R. (2013). Perception of material and shape of impacted everyday objects. In *Proceedings of the 10th International Symposium on Computer Music Multidisciplinary Research (CMMR 2013)*.

Rath, M. and Rocchesso, D. (2005). Continuous sonic feedback from a rolling ball. *IEEE Multimedia*, 12(2) :60–69.

Repp, B. and Su, Y. (2013). Sensorimotor synchronization : A review of recent research (2006–2012). *Psychonomic Bulletin & Review*, 20 :403–452.

Risset, J.-C. and Wessel, D. L. (1999). *The psychology of music*, chapter Exploration of timbre by analysis and synthesis, pages 113–169. Cognition and Perception. Academic Press, second edition.

Rosch, E. (1973). Natural categories. *Cognitive psychology*, 4(3) :328–350.

Rozé, J., Aramaki, M., Bourdin, C., Chadefaux, D., Dufrenne, M., Kronland-Martinet, R., Voinier, T., and Ystad, S. (2015a). Exploring the influence of cellists' postural movements on musical expressivity. In *Proceedings of the International Conference on the Multimodal Experience of Music*, Sheffield, UK.

Rozé, J., Aramaki, M., Kronland-Martinet, R., Voinier, T., Bourdin, C., Chadefaux, D., Dufrenne, M., and Ystad, S. (2015b). Assessing the influence of constraints on cellists' postural displacements and musical expressivity. In *Proceedings of the 11th International Symposium on*

Computer Music Multidisciplinary Research (CMMR) "Music, Mind and Embodiment", Plymouth, UK.

Rozé, J., Aramaki, M., Kronland-Martinet, R., Voinier, T., Bourdin, C., Chadefaux, D., Dufrenne, M., and Ystad, S. (2016). Exploring the effects of constraints on the cellist's postural displacements and their musical expressivity. In *Music, Mind and Embodiment, Post-proceedings of 11th International Symposium on Computer Music Modeling and Retrieval (CMMR 2015)*. Springer Berlin Heidelberg.

Schaeffer, P. (1966). *Traité des objets musicaux*. Ed. du Seuil.

Schaffert, N., Mattes, K., and Effenberg, A. (2010). A sound design for acoustic feedback in elite sports. In *Auditory Display*, Lecture Notes in Computer Science. Springer Verlag.

Sirdey, A., Derrien, O., Kronland-Martinet, R., and Aramaki, M. (2011). Modal analysis of impact sounds with ESPRIT in Gabor transforms. In *Proceedings of the Digital Audio Effects Conference : DAFx11*, pages 387–392, Paris, France.

Sirdey, A., Derrien, O., Kronland-Martinet, R., and Aramaki, M. (2012). ESPRIT in Gabor frames. In *Proceedings of the 45th Conference of Audio Engineering Society*, pages 305–313, Helsinki, Finlande.

Soto-Faraco, S., Kingstone, A., and Spence, C. (2003). Multisensory contributions to the perception of motion. *Neuropsychologia*, 41 :1847–1862.

Theurel, A., Frileux, S., Hatwell, Y., and Gentaz, E. (2012). The haptic recognition of geometrical shapes in congenitally blind and blindfolded adolescents : is there a haptic prototype effect. *PloS One*, 7(6).

Thoret, E. (2014). *Caractérisation acoustique des relations entre les mouvements biologiques et la perception sonore : application au contrôle de la synthèse et à l'apprentissage de gestes*. PhD thesis, Aix-Marseille Université.

Thoret, E., Aramaki, M., Bourdin, C., Bringoux, L., Kronland-Martinet, R., and Ystad, S. (2013a). Synchronizing gestures with friction sounds : Work in progress. In *Proceedings of the 10th International Symposium on Computer Music Multidisciplinary Research (CMMR 2013) "Sound, Music and Motion"*, Marseille, France.

Thoret, E., Aramaki, M., Bringoux, L., Ystad, S., and Kronland-Martinet, R. (2016a). Seeing circles and drawing ellipses : when sound biases reproduction of visual motion. *Plos One*, 11(4).

Thoret, E., Aramaki, M., Gondre, C., Kronland-Martinet, R., and Ystad, S. (2013b). Controlling a non linear friction model for evocative sound synthesis applications. In *Proceedings of the 16th International Conference on Digital Audio Effects (DAFx-13)*, Maynooth, Ireland.

Thoret, E., Aramaki, M., Gondre, C., Ystad, S., and Kronland-Martinet, R. (2016b). Eluding the physical constraints in a nonlinear interaction sound synthesis model for gesture guidance. *Applied Sciences*, 6(7) :192.

Thoret, E., Aramaki, M., Kronland-Martinet, R., Velay, J., and Ystad, S. (2012a). From shape to sound : Sonification of two dimensional curves by reenaction of biological movements. In *Proceedings of the 9th International Symposium on Computer Music Modelling and Retrieval (CMMR 2012) - Music & Emotions*, pages 278–285, Londres, Royaume-Uni.

Thoret, E., Aramaki, M., Kronland-Martinet, R., Velay, J., and Ystad, S. (2012b). Sonifying drawings : characterization of perceptual attributes of sounds produced by human gestures. In *Proceedings of the 11th Congrès Français d'Acoustique*, pages 1089–1094, Nantes, France.

Thoret, E., Aramaki, M., Kronland-Martinet, R., Velay, J., and Ystad, S. (2014). From sound to shape : Auditory perception of drawing movements. *Journal of Experimental Psychology : Human Perception and Performance*, 40 :983–994.

Thoret, E., Aramaki, M., Kronland-Martinet, R., and Ystad, S. (2013c). *Post-proceedings 9th International Symposium on Computer Music Modeling and Retrieval (CMMR 2012)*, chapter Reenacting Sensorimotor Features of Drawing Movements from Friction Sounds. Number 7900 in Lecture Notes in Computer Science. Springer Berlin Heidelberg.

Trémeau, F., Antonius, D., Cacioppo, J., Ziwich, R., Jalbrzikowski, M., Saccente, E., Silipo, G., Butler, P., and Javitt, D. (2009). In support of bleuler : objective evidence for increased affective ambivalence in schizophrenia based upon evocative testing. *Schizophrenia Research*, 107(2-3) :223–231.

van den Doel, K., Pai, D. K., Adam, T., Kortchmar, L., and Pichora-Fuller, K. (2002). Measurements of perceptual quality of contact sound models. In *Proceedings of the 2002 International Conference on Auditory Display*, Kyoto, Japan.

Van Petten, C. and Rheinfelder, H. (1995). Conceptual relationships between spoken words and environmental sounds : Event-related brain potential measures. *Neuropsychologia*, 33(4) :485–508.

Verron, C., Aramaki, M., Gonot, A., Scotti, T., Rakovec, C., Mingasson, A., and Kronland-Martinet, R. (2013). Event-driven interactive solid sound synthesis. In *Proceedings of the*

10th International Symposium on Computer Music Multidisciplinary Research (CMMR 2013) "*Sound, Music and Motion*", pages 960–966, Marseille, France.

Verron, C., Aramaki, M., Kronland-Martinet, R., and Pallone, G. (2008a). Spatialized additive synthesis. In *Acoustics'08*, Paris, France.

Verron, C., Aramaki, M., Kronland-Martinet, R., and Pallone, G. (2008b). Spatialized additive synthesis of environmental sounds. In *Proceedings of the 125th Convention of Audio Engineering Society*, San Francisco, USA. Convention Paper 7509, ISBN 978-0-937803-65-3, 2-6 Octobre 2008.

Verron, C., Aramaki, M., Kronland-Martinet, R., and Pallone, G. (2009a). Analysis/synthesis and spatialization of noisy environmental sounds. In *Proceedings of the 15th International Conference on Auditory Display (ICAD 2009)*, pages 36–40, Copenhague, Danemark. 18-21 Mai 2009.

Verron, C., Aramaki, M., Kronland-Martinet, R., and Pallone, G. (2010a). A 3d immersive synthesizer for environmental sounds. *IEEE Transactions on Audio, Speech, and Language Processing*, 18(6) :1550–1561.

Verron, C., Aramaki, M., Kronland-Martinet, R., and Pallone, G. (2010b). Contrôle intuitif d'un synthétiseur d'environnements sonores spatialisés. In *Proceedings of the 10th Congrès Français d'Acoustique*, Lyon, France.

Verron, C., Aramaki, M., Kronland-Martinet, R., and Pallone, G. (2010c). Spatialized synthesis of noisy environmental sounds. In Ystad, Aramaki, Kronland-Martinet, and Jensen, editors, *Auditory Display*, volume 5954 of *Lecture Notes in Computer Science*, pages 392–407. Springer-Verlag Berlin Heidelberg.

Verron, C., Kronland-Martinet, R., Pallone, G., and Aramaki, M. (2007). A spatialized additive synthesizer. In Schubert, E., Buckley, K., Eliott, R., Koboroff, B., Chen, J., and Stevens, C., editors, *Proceedings of the inaugural International Conference Music Communication Science (ICoMCS)*, Sydney, Australie. 5-7 Décembre 2007.

Verron, C., Pallone, G., Aramaki, M., and Kronland-Martinet, R. (2009b). Controlling a spatialized environmental sound synthesizer. In *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 321–324, New Paltz, NY. 18-21 Octobre 2009.

Viviani, P., Redolfi, M., and Baud-Bovy, G. (1997). Perceiving and tracking kinaesthetic stimuli : further evidence for motor-perceptual interactions. *Journal of Experimental Psychology : Human Perception and Performance*, 23 :1232–1252.

Viviani, P. and Stucchi, N. (1992). Biological movements look uniform : Evidence of motor-perceptual interactions. *Journal of Experimental Psychology : Human Perception and Performance*, 18 :603–623.

Viviani, P. and Terzuolo, C. (1982). Trajectory determines movement dynamics. *Neuroscience*, 7(2) :431–437.

Vogt, K., Pirro, D., Kobenz, I., H́oldrich, R., and Eckel, G. (2010). Physiosonic - evaluated movement sonification as auditory feedback in physiotherapy. In *Auditory Display*, Lecture Notes in Computer Science. Springer Verlag.

Wanderley, M. and Battier, M., editors (2000). *Trends in Gestural Control of Music*. IRCAM - Centre Pompidou.

Wersényi, G. (2010). Auditory representations of a graphical user interface for a better human-computer interaction. In *Auditory Display*, Lecture Notes in Computer Science. Springer Verlag.

Wessel, D. (1976). Perceptually based controls for additive synthesis. In *Proceedings of the International Computer Music Conference*.

Wessel, D. L. (1979). Timbre space as a musical control structure. *Computer Music Journal*, 3(2) :45–52.

Sixième partie

Sélection des publications principales

Cette partie du document regroupe une sélection de mes publications principales. Cette sélection reflète l'aspect pluridisciplinaire des recherches menées à travers un spectre de revues couvrant différents domaines allant de la physique, au traitement du signal jusqu'à la perception et la cognition. La dernière publication sélectionnée concerne un chapitre d'ouvrage dans "Guide to Brain-Computer Music Interfacing" (Ed. Springer). Elle fait l'objet des premières réflexions épistémologiques sur la démarche scientifique adoptée pour la détermination d'invariants morphologiques sonores et sur l'implication des paradigmes théoriques sous-jacents.

- M. Aramaki, H. Baillères, L. Brancheriau, R. Kronland-Martinet et S. Ystad. Sound quality assessment of wood for xylophone bars. **The Journal of the Acoustical Society of America**, 121(4) : 2407–2420, 2007.
- M. Aramaki, M. Besson, R. Kronland-Martinet et S. Ystad. Controlling the perceived material in an impact sound synthesizer. **IEEE Transactions on Audio, Speech, and Language Processing**, 19 (2) :301–314, 2011.
- E. Thoret, M. Aramaki, L. Bringoux, S. Ystad et R. Kronland-Martinet. Seeing circles and drawing ellipses : when sound biases reproduction of visual motion. **Plos One**, 11(4), 2016.
- S. Conan, E. Thoret, M. Aramaki, O. Derrien, C. Gondre, R. Kronland-Martinet et S. Ystad. An intuitive synthesizer of continuous interaction sounds : Rubbing, scratching and rolling. **Computer Music Journal**, 38(4) :24–37, 2014.
- E. Thoret, M. Aramaki, R. Kronland-Martinet, J.L. Velay et S. Ystad. From sound to shape : Auditory perception of drawing movements. **Journal of Experimental Psychology : Human Perception and Performance**, 40 :983–994, 2014.
- M. Aramaki, C. Marie, R. Kronland-Martinet, S. Ystad et M. Besson. Sound categorization and conceptual priming for nonlinguistic and linguistic sounds. **Journal of Cognitive Neuroscience**, 22(11) :2555–2569, November 2010.
- J.A. Micoulaud-Franchi, M. Aramaki, A. Merer, M. Cermolacce, S. Ystad, R. Kronland-Martinet, J. Naudin et J. Vion-Dury. Towards an exploration of feeling of strangeness in schizophrenia : Perspectives on acousmatic and everyday listening. **Journal of Abnormal Psychology**, 121(3) :628–640, 2012.
- M. Aramaki, R. Kronland-Martinet, S. Ystad, J.A. Micoulaud-Franchi and J. Vion-Dury. Prospective view on sound synthesis BCI control in light of two paradigms of cognitive neuroscience. In E. Miranda et J. Castet, éditeurs : Guide to Brain-Computer Music Interfacing, pages 61–87. Springer Berlin Heidelberg, 2014.

Sound quality assessment of wood for xylophone bars

Mitsuko Aramaki[a)]
CNRS Laboratoire de Mécanique et d'Acoustique 31, chemin Joseph Aiguier 13402 Marseille Cedex 20, France

Henri Baillères and Loïc Brancheriau
CIRAD-Forêt, TA 10/16, avenue Agropolis, 34398 Montpellier Cedex 5, France

Richard Kronland-Martinet and Sølvi Ystad
CNRS, Laboratoire de Mécanique et d'Acoustique 31, chemin Joseph Aiguier 13402 Marseille Cedex 20, France

(Received 15 March 2006; revised 22 January 2007; accepted 22 January 2007)

Xylophone sounds produced by striking wooden bars with a mallet are strongly influenced by the mechanical properties of the wood species chosen by the xylophone maker. In this paper, we address the relationship between the sound quality based on the timbre attribute of impacted wooden bars and the physical parameters characterizing wood species. For this, a methodology is proposed that associates an analysis-synthesis process and a perceptual classification test. Sounds generated by impacting 59 wooden bars of different species but with the same geometry were recorded and classified by a renowned instrument maker. The sounds were further digitally processed and adjusted to the same pitch before being once again classified. The processing is based on a physical model ensuring the main characteristics of the wood are preserved during the sound transformation. Statistical analysis of both classifications showed the influence of the pitch in the xylophone maker judgement and pointed out the importance of two timbre descriptors: the frequency-dependent damping and the spectral bandwidth. These descriptors are linked with physical and anatomical characteristics of wood species, providing new clues in the choice of attractive wood species from a musical point of view. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2697154]

PACS number(s): 43.75.Kk, 43.66.Jh, 43.60.Uv [NFH] Pages: 2407–2420

I. INTRODUCTION

The mechanical and anatomical properties of woods are of importance for the sound quality of musical instruments. Yet, depending on the role of the wooden elements, these properties may differ. Xylophone sounds are produced by striking wooden bars with a mallet, and thus the mechanical properties of the wood are important. This study is the first step towards understanding what makes the sound of an impacted wooden bar attractive for xylophone makers from a musical point of view. For this purpose, we recorded sounds from a wide variety of wood species to compare their sound quality and relate it to the wood properties. An original methodology is proposed that associates analysis-synthesis processes and perceptual classification analysis. Perceptual classification was performed by a renowned instrument maker.

The xylophone maker community agrees on the choice of wood species. This choice is driven by the sound quality, but other nonacoustically relevant properties are considered as well (e.g., robustness; esthetic aspects). The wood species most used in xylophone manufacturing is *Dalbergia* sp. Several authors have sought to determine which physical characteristics are of importance for the generated sound. In particular, Holz (1996) concluded that an "ideal" xylophone wood bar is characterized by a specific value range of density, Young modulus, and damping factors. Ono and Norimoto (1983) demonstrated that samples of spruce wood (*Picea excelsa, P. glehnii, P. sitchensis*)—considered a suitable material for soundboards—all had a high sound velocity and low longitudinal damping coefficient as compared to other softwoods. The cell-wall structure may account for this phenomenon. Internal friction and the longitudinal modulus of elasticity are markedly affected by the microfibril angle in the S2 tracheid cell layer, but this general trend does not apply to all species. For instance, pernambuco (*Guilandina echinata* Spreng.), traditionally used for making violin bows, has an exceptionally low damping coefficient relative to other hardwoods and softwoods with the same specific modulus (Bucur, 1995; Matsunaga *et al.*, 1996; Sugiyama *et al.*, 1994). This feature has been explained by the abundance of extractives in this species (Matsunaga and Minato, 1998). Obataya *et al.*(1999) confirmed the importance of extractives for the rigidity and damping qualities of reed materials. Matsunaga *et al.* (1999) reduced the damping coefficient of spruce wood by impregnating samples with extractives of pernambuco (*Guilandina echinata* Spreng.). The high sound quality conditions are met by the wood species commonly used by xylophone makers (like *Dalbergia* sp.), but other tropical woods may serve. We propose to focus on the perceptual properties of impacted wood bars as the basis for pointing out woods suitable for xylophone manufacturing. Several studies using natural or synthetic sounds have been conducted to point out auditory clues associated with geom-

[a)]Author to whom correspondence should be addressed. Electronic mail: aramaki@lma.cnrs-mrs.fr

etry and material properties of vibrating objects (Avanzini and Rocchesso, 2001; Giordano and McAdams, 2006; Lutfi and Oh, 1997; Klatzky *et al.*, 2000; McAdams *et al.*, 2004). These studies revealed the existence of perceptual clues allowing the source of the impact sound to be identified merely by listening. In particular, the perception of material correlated mainly with the internal friction (related to the damping factors of the spectral components) as theoretically shown by Wildes and Richards (1988). Nevertheless, it has not been determined whether the perceptual clues highlighted in the distinction of different materials are those used to establish the subjective classification of different species of wood.

The perceptual differences reported in the literature are linked with subtle changes in timbre, defined as "the perceptual attribute that distinguishes two tones of equal, pitch, loudness, and duration" (ANSI, 1973). This definition points out the importance of comparing sounds with similar loudness, duration, and pitch. Concerning loudness and duration, the sounds of interest can easily be adjusted in intensity by listening, and they have about the same duration since they correspond to the very narrow category of impacted wooden bars. Concerning pitch, the bars do not have the same values because the pitch depends on the physical characteristics of the wood, i.e., essentially of the Young modulus and the mass density. To tune the sounds to the same pitch, we propose to digitally process the sounds recorded on bars of equal length. Synthesis models can be used for this purpose, allowing virtual tuning by altering the synthesis parameters. Such an approach combining sound synthesis and perceptual analysis has already been proposed. Most of the proposed models are based on the physics of vibrating structures, leading to a modal approach of the synthesis process (Adrien, 1991; Avanzini and Rocchesso, 2001) or to a numerical method of computation (Bork, 1995; Chaigne and Doutaut, 1997; Doutaut *et al.*, 1998). Yet, although these models lead to realistic sounds, they do not easily allow for an analysis-synthesis process implicating the generation of a synthetic sound perceptually similar to an original one. To overcome this drawback, we propose an additive synthesis model based on the physics of vibrating bars, the parameters of which can be estimated from the analysis of natural sounds.

The paper is organized as follows: in Sec. II, we discuss the design of an experimental sound data bank obtained by striking 59 wooden bars made of different woods carefully selected and stabilized in a climatic chamber. In Sec. III, we then address the issue of digitally tuning the sounds without changing the intrinsic characteristics of the wood species. This sound manipulation provided a tuned sound data bank in which each sound was associated with a set of descriptors estimated from both physical experiments and signal analysis. The experimental protocol is described in Sec. IV. It consists of the classification carried by a professional instrument maker. The classification was performed with both the original and the tuned data banks to better understand the influence of pitch on the classification. These results are discussed in Sec. VII, leading to preliminary conclusions that agree with most of the knowledge and usage in both wood mechanics, xylophone manufacturing, and sound perception.

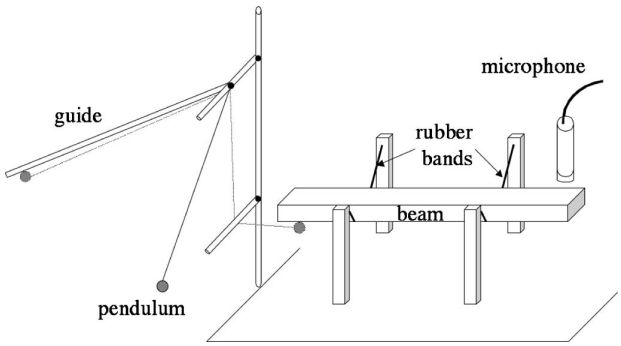


FIG. 1. Experimental setup used to strike the wood samples and record the impact sounds. The setup was placed in an anechoic room.

II. DESIGN OF AN EXPERIMENTAL SOUND DATA BANK

a. Choice of wood species. Most percussive instruments based on wooden bars are made of specific species (for example, *Dalbergia* sp. or *Pterocarpus* sp.). In this experiment, we used tropical and subtropical species, most of which were unknown to instrument makers. A set of 59 species presenting a large variety of densities (from 206 to 1277 kg/m^3) were chosen from the huge collection (about 8000) of the CIRAD (Centre de coopération Internationale en Recherche Agronomique pour le Développement, Montpellier, France). Their anatomical and physical characteristics have been intensely studied and are well described. The name and density of each species are in Table III.

b. Manufacturing wooden bars. Both geometry and boundary conditions govern the vibration of bars. By considering bars with the same geometry and boundary conditions, sounds can be compared to determine the intrinsic quality of the species. Hence, a set of bars was made according to the instrument maker recommendations. The bars were manufactured to be as prismatic as possible, with dimensions $L=350\ mm \times W=45\ mm \times T=20\ mm$, without singularities and cut in the grain direction. We assume that the growth rings are parallel to the tangential wood direction and that their curvature is negligible. The longitudinal direction is collinear to the longitudinal axis of the bars. The bars were stabilized in controlled conditions.

c. Recording of impact sounds under anechoic conditions. An experimental setup was designed that combines an easy way to generate sounds with a relative precision ensuring the repeatability of the measurements, as shown in Fig. 1. In this way, impact excitation was similar for all the impacted bars. Moreover, to minimize the sound perturbations due to the environment, the measurements took place in an anechoic room.

The bar was placed on two rubber bands, ensuring free-free-type boundary conditions. The rubbers minimized perturbations due to suspension (see, for example, Blay *et al.*, 1971 for more details). Bars were struck with a small steel pendulum. The ball on the string was released from a constrained initial position (guide), and after the string wrapped around a fixed rod, the ball struck the bar from underneath. The robustness of this simple procedure showed the radiated

sounds were reproducible: the determination error was less than 0.1% for the fundamental frequency and 4.3% for the damping coefficient of the first mode (Brancheriau *et al.*, 2006a). To ensure broad spectral excitation, the ball was chosen to generate a sufficiently short pendulum/bar contact (to be as close as possible to an ideal Dirac source). The excitation spectrum is given by the Fourier transform of the impact force, so that the shorter the impact, the broader the spectrum excitation. For that, a steel ball was used since the modulus of elasticity of steel is much larger than that of wood (the ratio is about 200). This setup makes contact duration between the ball and the bar short (Graff, 1975). This duration was shortened because the impact point was underneath the bar, maximizing the reversion force. After several experiments, a good compromise between speed, short duration, and lack of deformation of the material was obtained with a steel ball of 12 g and a 14 mm diameter, tightened by a 30-cm-long string. The impact point played an important role in the generation of sounds. To prevent the first modes from vanishing, the bar was struck close to one of its extremities (at 1 cm), allowing high frequency modes to develop. An omni-directional microphone (Neumann KM183mt) was placed in the close sound field at the opposite end of the impact location to measure the sound-radiated pressure. This configuration obviates the contribution of the spectral peak generated by the ball, peak which was at about 10 kHz. The sounds were digitally recorded at 48 kHz sampling frequency.

d. Signal characteristics. Figure 2 shows the temporal signal, the spectral representation, and the time-frequency representation of a typical sound obtained experimentally. The temporal signals are characterized by a short onset and a fast decay. Consequently, their durations generally do not exceed 1 s. Their spectra are composed of emergent resonances that do not overlap much. As shown by the time-frequency representation, the damping of these spectral components is frequency dependent, the high frequency components being more heavily damped than the low frequency ones.

III. DESIGN OF TUNED SOUND DATA BANK FOR TIMBRE STUDY

To facilitate comparison of the timbre of sounds generated striking different wood species, their pitch was equalized. In practice, this could have been possible using the same procedure adopted by percussive instrument makers, where the bar geometry is modified removing some substance around the center of the bar to be tuned (Fletcher and Rossing, 1998). This approach comes, however, with the risk of making irreversible mistakes, for example, removing an excessive amount of wood. As an alternative, we propose to digitally tune the pitch of sounds generated striking bars of equal length. Such an approach relies on the plausible assumption that the pitch of our recorded signals is primarily determined by the frequency of the first vibrational mode. In particular, we use a sound synthesis model which allows for sound transformations that are accurate relative to the physical phenomena, as compared to other signal processing approaches such as pitch shifting.

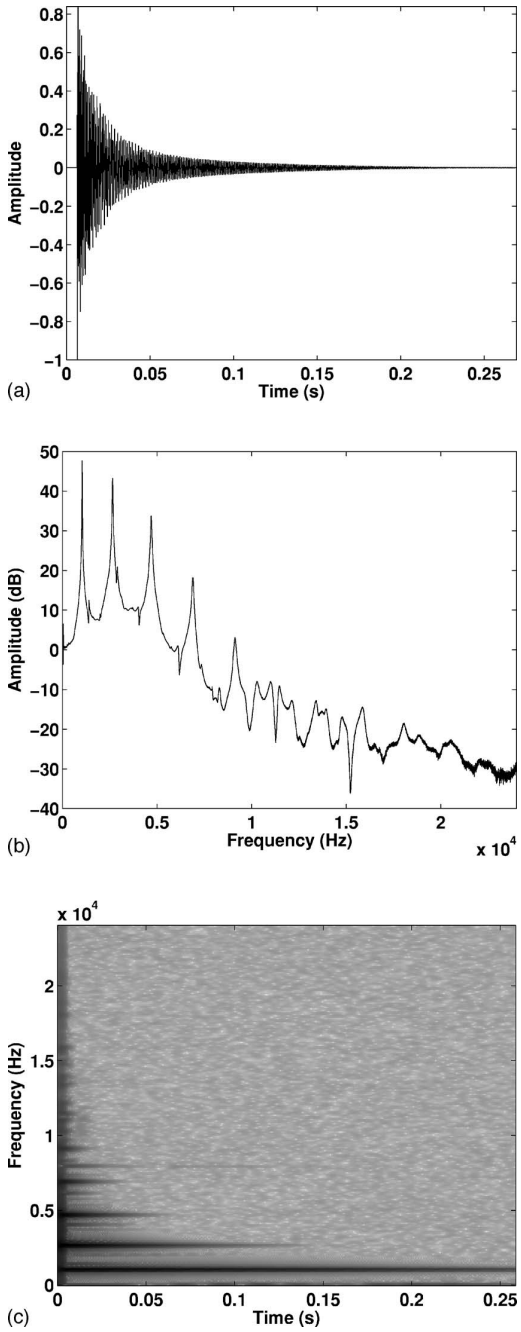


FIG. 2. (a) Wave form, (b) spectral representation, and (c) spectrogram (amplitude in logarithmic scale) of a typical sound obtained by impacting a wooden bar.

A. Synthesis model based on physical approach

To tune the sounds, we propose to use an additive synthesis model. This model simulates the main characteristics of the vibrations produced by an impacted bar to exhibit the principal properties of the radiated sound.

1. Simplified mechanical model

Numerous mechanical models of bar vibrations are available in the literature, but the relevant information can be

pointed out using a simple model based on assumptions that are coherent with our experimental design. According to the manufacturing of the bars, one can assume that the fiber orientation follows the axis of the bar and that the ratio length/width is large. Consequently, one can neglect the anisotropy property of the wood and the contribution of the longitudinal and torsional modes (which are few, weak, and of little influence on the radiated sound). These assumptions allow for the consideration of a one-dimensional mechanical model depending only on the longitudinal Young modulus. Such a model can be described by the well-known Euler-Bernoulli equation

$$EI\frac{\partial^4 y(x,t)}{\partial x^4} + \rho S\frac{\partial^2 y(x,t)}{\partial t^2} = 0, \tag{1}$$

where E is the longitudinal Young modulus, I the quadratic moment, ρ the mass density, and S the cross section area. The general solution of the equation is given by

$$y(x,t) = \sum_n Y_n(x)e^{i\gamma_n t} \tag{2}$$

with

$$Y_n(x) = A\cosh(k_n x) + B\sinh(k_n x) + C\cos(k_n x) + D\sin(k_n x). \tag{3}$$

By injecting Eq. (2) and Eq. (3) into the Eq. (1), one obtains

$$\gamma_n = \pm\sqrt{\frac{EI}{\rho S}}k_n^2. \tag{4}$$

Our experimental setup corresponds to free-free boundary conditions written

$$\frac{\partial^2 Y(0)}{\partial x^2} = \frac{\partial^2 Y(L)}{\partial x^2} = \frac{\partial^3 Y(0)}{\partial x^3} = \frac{\partial^3 Y(L)}{\partial x^3} = 0$$

leading to

$$k_n = (2n+1)\frac{\pi}{2L}. \tag{5}$$

To take into account viscoelastic phenomena, E is considered as complex valued, see, for example (Valette and Cuesta, 1993)

$$E = E_d(1 + i\eta), \tag{6}$$

where E_d is the dynamical Young modulus, and η a dimensionless material loss factor. By injecting relations (5) and (6) into relation (4) and assuming that $\eta \ll 1$, one obtains the following important expressions:

$$\gamma_n = \omega_n + i\alpha_n \tag{7}$$

with

$$\begin{cases} \omega_n \approx \sqrt{\frac{E_d I}{\rho S}}(2n+1)^2\frac{\pi^2}{4L^2} \\ \alpha_n \approx \frac{\eta}{2}\omega_n \end{cases}. \tag{8}$$

Thus, one can rewrite the relation (2):

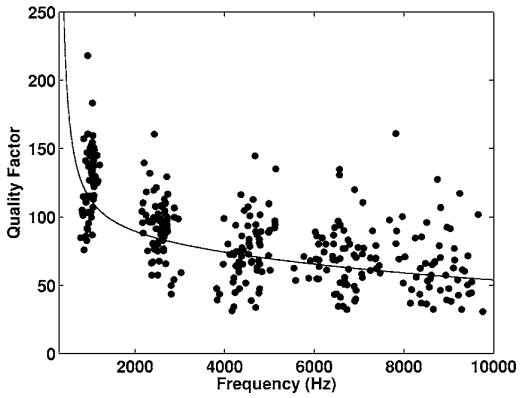


FIG. 3. Quality factor Q_n estimated on original sounds. The Q_n values are fitted, in a least squares sense, by a rational function (black curve) corresponding to Eq. (11).

$$y(x,t) = \sum_n Y_n(x)e^{i\omega_n t}e^{-\alpha_n t}. \tag{9}$$

It is accepted (Chaigne and Doutaut, 1997; McAdams *et al.*, 2004; Ono and Norimato, 1985) that the damping factors in case of wooden bars are described by a parabolic form:

$$\alpha(f) = a_0 + a_2 f^2 \tag{10}$$

where the constants a_0 and a_2 depend on the wood species. This corresponds to a quality factor Q_n given by

$$Q_n = \frac{\pi f_n}{\alpha_n} = \frac{\pi f_n}{a_0 + a_2 f_n^2}. \tag{11}$$

This behavior was experimentally verified, as shown in Fig. 3.

These expressions show that the vibrations of the bar, which are correlated with the radiated sound pressure, can be described by a sum of elementary components consisting of exponentially damped monochromatic signals. The frequency of these elementary components is inversely proportional to the square of the length of the bar, and their damping is proportional to the square of the frequency.

2. Additive synthesis model

The synthesis model aims at simulating the analytical solutions written in Eq. (9), which are expressed as a sum of exponentially damped sinusoids

$$s(x,t) = \theta(t)\sum_{n=1}^{N} A_n(x)\sin(\omega_n t)e^{-\alpha_n t}, \tag{12}$$

where N is the number of components, $\theta(t)$ the Heaviside function, A_n the amplitude, ω_n the frequency and α_n the damping coefficient of the nth component. The choice of either sine or cosine functions has no perceptual influence on the generated sounds but sine functions are often used in sound synthesis since they avoid discontinuities in the signal at $t=0$. Hence, the signal measured at a fixed location is considered to be well represented by the expression (12). Its spectral representation is given by

$$S(\omega) = \sum_{n=1}^{N} \frac{A_n}{2i}\left(\frac{1}{\alpha_n + i(\omega - \omega_n)} - \frac{1}{\alpha_n + i(\omega + \omega_n)}\right)$$

and the z transform by

$$S(z) = \sum_{n=1}^{N} \frac{A_n}{2i}\left(\frac{1}{1 - e^{(i\omega_n - \alpha_n)}z^{-1}} - \frac{1}{1 - e^{(-i\omega_n - \alpha_n)}z^{-1}}\right).$$

B. Estimation of synthesis parameters

Before the tuning process, the recorded sounds described in Sec. II are equalized in loudness, analyzed, and then resynthesized with the synthesis model described above. The loudness was equalized by listening tests. For that, the synthesis parameters are directly estimated from the analysis of the recorded sounds. The estimation of the parameters defining the sound is obtained by fitting the recorded signal with the expression given in relation (12). To do so, we used a signal processing approach that consists of identifying the parameters of a linear filter by auto regressive and moving average (ARMA) analysis. We model the original signal as the output of a generic linear filter whose z transform is written

$$H(z) = \frac{\sum_{m=0}^{M} a_m z^{-m}}{1 + \sum_{n=1}^{N} b_n z^{-n}} = a_0 z^{N-M} \frac{\prod_{m=1}^{M} (z - z_{0m})}{\prod_{n=1}^{N} (z - z_{pn})},$$

where z_{0m} are the zeros and z_{pn} are the poles of the system. Only the most prominent spectral components were modeled by $H(z)$. These spectral components were determined within a 50 dB amplitude dynamic, the reference being the amplitude of the most prominent spectral peak. Hence, the number of poles N and zeros M of the linear ARMA filter is determined by the number of spectral components taken into account. The coefficients a_m and b_n are estimated using classical techniques such as Steiglitz-McBride (Steiglitz and McBride, 1965). The synthesis parameters corresponding to the amplitudes, frequencies, and damping coefficients of the spectral components are thus determined:

$$\begin{cases} A_n = |H(z_{pn})|, \\ \omega_n = \arg(z_{pn}) f_s, \\ \alpha_n = \log(|z_{pn}|) f_s, \end{cases} \tag{13}$$

where f_s is the sampling frequency. In addition to the synthesis model described above, we have taken into account the attack time. Actually, even though the rising time of the sounds is very short, it does influence the perception of the sounds. These rising times were estimated on the original sounds and were reproduced by multiplying the beginning of the synthetic signal by an adequate linear function. Synthesis sounds were evaluated by informal listening tests confirming that their original sound qualities were preserved. The synthesis quality was further confirmed by results from the professional instrument maker showing a similar classification of original and synthetic sounds (classifications Cl and C2, see Sec. VI A 1).

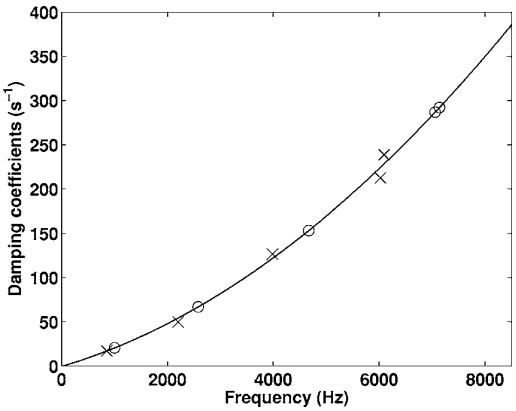


FIG. 4. The damping coefficients of the original sound (×) are fitted by a parabolic function (solid curve). The damping coefficients of the tuned sound (◦) are determined according to this damping law.

C. Tuning the sounds

The processing of tuning the sounds at the same pitch was based on some assumptions specific to the investigated stimulus set and consistent with the vibratory behavior of the bar. For the kind of sounds we are dealing with (impacted wooden bars), we assume the pitch to be related to the frequency of the first vibration mode, which is correlated with the length of the bar [cf. Eq. (8)]. Actually, if the length L changes to βL, then ω_n changes to ω_n/β^2. As a consequence, a change in pitch corresponds to a dilation of the frequency components. These assumptions made it possible to virtually equalize the pitches of the recorded bank of sounds. To minimize the pitch deviation, the whole set of sounds was tuned by transposing the fundamental frequencies to 1002 Hz, which is the mean fundamental frequency of all the sounds. The amplitude of the spectral components was kept unchanged by the tuning process. Once again, no precise listening test was performed, but our colleagues found the synthesis sounds preserved the specificity of the material.

According to the discussion in III A 1, the damping is proportional to the square of the frequency. Thus, from the expression (10), a damping law can be defined by a parabolic function that can be written in a general form:

$$\alpha(\omega) = D_A\omega^2 + D_B\omega + D_C. \tag{14}$$

As a consequence, when the pitch is changed, the damping coefficient of each tuned frequency component has to be evaluated according to the damping law measured on the original sound (cf. Fig. 4).

Figure 5 shows the comparison between the spectrum of a measured signal and the spectrum of a tuned signal after the resynthesis process. The entire sound data bank is available at http://www.lma.cnrs-mrs.fr/~kronland/JASA_Xylophone/sounds.html.

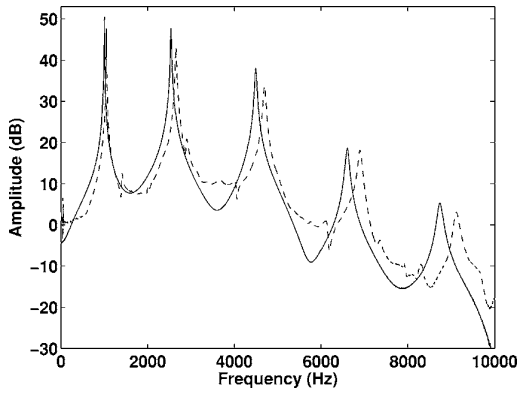


FIG. 5. Comparison between a spectrum of a measured signal (dashed trace) and the spectrum of the associated tuned signal (solid trace).

IV. EXPERIMENTAL PROTOCOL

Sounds from different wooden bars were evaluated focusing on the perceived musical quality of the wood samples. The participant was placed in front of a computer screen on which the sounds (all represented as identical crosses) were randomly distributed. The participant was asked to place the sounds on a bidimensional computer display. In particular, he was told that the horizontal dimension of the display represented an axis of musical quality so that sounds judged as having the worst/best quality were to be placed on the leftmost/rightmost part of the display. The participant could listen to the sounds as often as he wanted by simply clicking on the cross. The tests were carried on a laptop Macintosh equipped with a Sony MDR CD550 headset.

For this study, one instrument maker specialized in xylophone manufacture carried the task. For a complete perceptual study, more participants would, of course, be needed. As a first step we aimed at presenting a new methodology for an interdisciplinary approach uniting instrument makers and specialists within acoustics, signal processing, and wood sciences.

Three tests were conducted using this experimental protocol. The instrument maker carried the classification with the original sounds (recorded sounds with different pitches), called C1 (Brancheriau *et al.*, 2006a; Brancheriau *et al.*, 2006b). A second classification, called C2, using the synthesized sounds (resynthesis of the original sounds with different pitches) was done two years later. The comparison of C1 and C2 allowed us to check the quality of the resynthesis as well as the reliability of our experimental participant. The third test (C3) was carried on the signals tuned to the same pitch. The xylophone maker was not aware of the synthetic nature of sounds in C2 and C3. In particular, he was told that, for classification C3, the same pieces of wood had been sculpted in order to tune the sounds to the same fundamental frequency. Classification C3 is presented in Table III.

V. DESCRIPTORS

A. Mechanical descriptors

The wood species used for this study have been intensively examined at CIRAD and their anatomical and physical characteristics are well known. Thus, the mechanical descriptors are defined by the mass density, ρ, the longitudinal modulus of elasticity, E_ℓ, and the transverse shear modulus, G_t. The descriptors E_ℓ and G_t can be calculated using Timoshenko's model and the Bordonné solutions (Brancheriau and Baillères, 2002). We have also considered the specific longitudinal modulus, E_ℓ/ρ, and the specific shear modulus, G_t/ρ.

B. Signal descriptors

To characterize the sounds from an acoustical point of view, we calculated the following timbre descriptors (Caclin *et al.*, 2005; McAdams *et al.*, 1995): attack time, AT (the way the energy rises during the onset of the sound), spectral bandwidth, SB (spectrum spread), spectral centroid, SCG (brightness), and spectral flux, SF (the way the sound vanishes).

The attack time, AT, a temporal descriptor, characterizes the signal onset and describes the time it takes for the signal to reach its maximum. It is generally estimated as the time it takes the signal to deploy its energy from 10% to 90% of the maximum. The spectral timbre descriptors characterize the organization of the spectral peaks resulting from the modal behavior of the bar vibration. One of the most well known is the spectral centroid, SCG, which is correlated with the subjective sensation of brightness (Beauchamps, 1982):

$$\mathrm{SCG} = \frac{\sum_k f(k)|\hat{s}(k)|}{\sum_k |\hat{s}(k)|}, \tag{15}$$

where $\hat{s}$ is the discrete Fourier transform of the signal $s(t)$ and f the frequency. The spectral bandwidth, SB, measures the spread of the spectral components around the spectral centroid and is defined as (Marozeau, de Cheveigné, McAdams and Winsberg, 2003)

$$\mathrm{SB} = \sqrt{\frac{\sum_k |\hat{s}(k)|(f(k) - \mathrm{SCG})^2}{\sum_k |\hat{s}(k)|}}. \tag{16}$$

Finally, the fourth classical timbre descriptor called the spectral flux, SF, is a spectro-temporal descriptor that measures the deformation of the spectrum with respect to time. In practice, the spectral flux is given by a mean value of the Pearson correlation calculated using the modulus of local spectral representations of the signal (McAdams *et al.*, 1995):

$$\mathrm{SF} = \frac{1}{N}\sum_{n=1}^{N} \frac{\langle s_n, s_{n-1}\rangle}{s_n^2 s_{n-1}^2}, \tag{17}$$

where N represents the number of frames, s_n the modulus of the local spectrum at the discrete time n, and $\langle,\rangle$ the discrete scalar product.

In addition to these well-known timbre descriptors, we propose to consider various acoustical parameters chosen as function of the specificities of the impact sounds, i.e., the

amplitude ratio between the first two frequency components of the sound, noted $A_{2/1}$, and the damping and the inharmonicity descriptors. The last two parameters are described below in more detail. The damping descriptor is defined from the Eq. (14) by the set of coefficients $\{D_A, D_B, D_C\}$ traducing the sound decrease. As the damping is the only parameter responsible for the variation of the spectral representation of the signal with respect to time, this descriptor is related to the spectral flux, SF. In addition, the damping coefficients α_1 and α_2 of components 1 and 2 have been included in the list of signal descriptors. The inharmonicity characterizes the relationship between the partials and the fundamental mode. This parameter is linked with the consonance, which is an important clue in the perceptual differentiation of sounds. For each spectral component, inharmonicity is defined by

$$I(n) = \frac{\omega_n}{\omega_0} - n. \tag{18}$$

From this expression, we propose an inharmonicity descriptor defined by a set of coefficients $\{I_A, I_B, I_C\}$ obtained by fitting $I(n)$ with a parabolic function, as suggested by the calculation $I(n)$ from Eq. (8):

$$I(n) = I_A n^2 + I_B n + I_C. \tag{19}$$

VI. RESULTS

Collected behavioral data could be considered as ordinal. Nevertheless, since the task consisted in placing the sounds on a quality axis "as a function of its musical quality," the relative position of the sounds integrates a notion of perceptual distance. Moreover, the classifications do not contain two sounds with the same position and do not show categories (see Table III). It was thus decided to consider the data as providing a quantitative estimate of perceived musical quality for the wood samples, the value associated with each species being given by its abscissa from 0 (worst quality) to 10 (best quality) on the quality axis. The main interest in using quantitative scales is the possibility of constructing an arithmetic model for perceived wood quality which can be easily used to estimate the musical quality of woods (and sounds) not considered in our experiments. All the statistical analyses were conducted with SPSS software (Release 11.0.0, LEAD Technologies).

A. Qualitative analysis—Choice of the variables

1. Resynthesis quality—Robustness of the classification

Only one participant performed the classifications on the basis of his professional skill, and his judgment of sound quality was used to build reference quality scales. The xylophone maker is thus considered as a "sensor" for measuring the acoustical wood quality. The raw classifications C1 and C2 were compared using the Wilcoxon signed rank test to evaluate the resynthesis quality of the model. Moreover, this comparison allowed us to evaluate the robustness of the xylophone maker classification. No particular distribution was assumed for the classifications. The Wilcoxon test is thus appropriate for comparing the distributions of the two classifications (C1, C2). The significance value of the Wilcoxon test ($p=0.624$) for (C1, C2) indicates that classification C1 equals classification C2. There was no significant difference in the xylophone maker responses between C1 and C2.

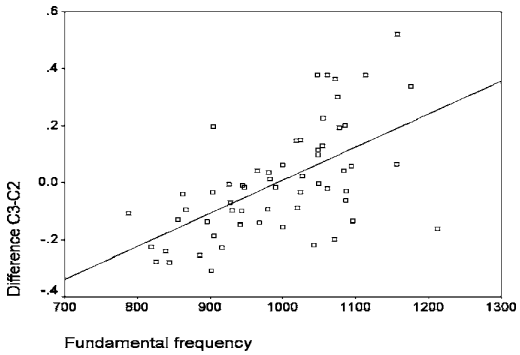


FIG. 6. Linear relationship between fundamental frequency and arithmetic difference C3-C2 ($R=0.59$, $N=59$).

2. Influence of the tuning process

The same Wilcoxon signed rank test was performed with classification C2 and classification C3 of tuned sounds. The hypothesis of equal distribution is rejected considering classifications C2 and C3. A significant difference between C2 and C3 ($p=0.001$) is due to the tuning process of sounds, which altered the sound perception of the xylophone maker. The arithmetic difference (C3-C2) was thus computed and related to the value of the fundamental frequency by using the Pearson correlation coefficient (Fig. 6). This coefficient value was found significant at the 1% level ($R=0.59$).

B. Quantitative analysis

1. Descriptor analysis

The 18 parameters presented in Table I were estimated for the tuned sounds and using standard mechanical calibrations. They are grouped into mechanical/physical descriptors and signal descriptors. In practice, for the spectral descriptors, the Fourier transform was estimated using a fast Fourier transform (FFT) algorithm. The length of the FFT was chosen so that it matches the longest sound, i.e., 2^{16} samples. For the SF calculation, the number of samples was 256 with an overlap of 156 samples. A Hamming window was used to minimize the ripples. Mechanical descriptors are linked with the intrinsic behavior of each sample but also linked with signal descriptors, as shown in Fig. 7. Indeed, the bivariate coefficients of determination matrix calculated on the basis of the 18 characteristic parameters revealed close collinearity between the parameters. Considering the strong relationship between the parameters, the statistical analyses were conducted by grouping the mechanical/physical descriptors and the signal descriptors in order to find those that best explain the classification C3.

A principal component analysis was thus conducted (Table II). Principal components analysis finds combinations of variables (components) that describe major trends in the data. This analysis generated a new set of parameters derived from the original set in which the new parameters (principal

TABLE I. Mechanical and signal descriptors computed from dynamic tests.

	No.	Variable	Signification
Mechanical descriptors	1	ρ	Mass density (kg/m^3)
	2	E_ℓ	Longitud. modulus of elasticity (MPa)
	3	G_t	Shear modulus (MPa)
	4	E_ℓ/ρ	Specific longitudinal modulus
	5	G_t/ρ	Specific shear modulus
Signal descriptors	6	$A_{2/1}$	Amplitude ratio of mode 2 and 1
	7	α_1	Temporal damping of mode 1 (s^{-1})
	8	α_2	Temporal damping of mode 2 (s^{-1})
	9	SCG	Spectral centroid (Hz)
	10	SB	Spectral bandwidth (Hz)
	11	SF	Spectral flux
	12	AT	Attack time (ms)
	13	D_A	Coefficient D_A of $\alpha(\omega)$
	14	D_B	Coefficient D_B of $\alpha(\omega)$
	15	D_C	Coefficient D_C of $\alpha(\omega)$
	16	I_A	Coefficient I_A of $I(n)$
	17	I_B	Coefficient I_B of $I(n)$
	18	I_C	Coefficient I_C of $I(n)$

components) were not correlated and closely represented the variability of the original set. Each original parameter was previously adjusted to zero mean and unit variance so that eigenvalues could be considered in choosing the main factors. In this case, the eigenvalues sum the number of variables, and eigenvalues can be interpreted as the number of original variables represented by each factor. The principal components selected thus corresponded to those of eigenvalue superior or equal to unity. Table II shows that six principal components accounted for 87% of all information contained in the 18 original parameters.

The relationships between original variables and principal components are presented in Figs. 8(a) and 8(b). These figures display the bivariate coefficient of determination between each principal component and each original parameter; the bivariate coefficient corresponds to the square loading coefficient in this analysis. The variance of the inharmonicity coefficients $\{I_A, I_B, I_C\}$ and the damping coefficients $\{D_A, D_B, D_C\}$ are captured by the first principal component and to a lesser degree by the third component [Fig. 8(a)]. The damping coefficients (α_1 and α_2), however, are mainly linked with the second component. This component is also linked with the amplitude ratio $A_{2/1}$ and with the timbre descriptors (SCG, SB, SF, AT). The variance of the mechanical/physical descriptors is scattered between all the principal components (parameter 1 is linked with PC1 and 2; parameter 2 with PC1 and 4; parameter 3 with PC3 and 5; parameter 4 with PC2, 3, and 4; and parameter 5 with PC3 and 5).

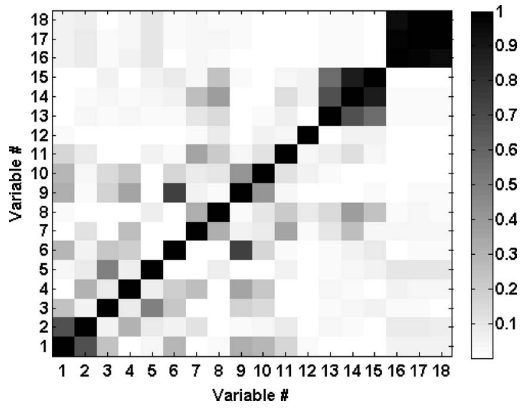


FIG. 7. Bivariate coefficients of determination for characteristic parameters (N=59).

2. Relationship between the descriptors and the acoustic classification of tuned sounds

a. Bivariate analysis. Figure 9 presents the results of bivariate analysis between characteristic parameters and classification C3. Assuming a linear relationship, the parameter α_1 (temporal damping of mode 1) appeared to be the best individual predictor with a R^2 value of 0.72. The second most significant predictor was the spectral flux, SF, with a R^2 value of 0.38. The other parameters were of minor importance considering classification C3. Note that the only mechanical parameter of interest was E_ℓ/ρ (specific longitudinal modulus) with a relatively low R^2 value of 0.25. Furthermore, the mass density, ρ, was not reflected in the acoustic classification (no significant R^2 value at the 1% level). Light woods and heavy woods were thus not differ-

TABLE II. Variance explained by the principal components (number of initial variables=18, number of samples=59).

Component	Eigen val.	% of Var.	Cumul. (%)
I	4.0	22.5	22.5
II	3.9	21.9	44.3
III	3.5	19.3	63.7
IV	1.8	10.1	73.8
V	1.2	6.7	80.5
VI	1.1	6.1	86.6

TABLE III. Botanical names of wood species ($N=59$), their density (kg/m^3), α_1 the temporal damping of mode 1(s^{-1}), SB the spectral bandwidth (Hz) and classification C3 by the xylophone maker (normalized scale from 0 to 10).

Botanical name	Density (kg/m^3)	α_1(s^{-1})	SB (Hz)	C3
Pericopsis elata Van Meeuw	680	21.76	2240	5.88
Scottellia klaineana Pierre	629	23.97	2659	6.38
Ongokea gore Pierre	842	26.07	2240	5.15
Humbertia madagascariensis Lamk.	1234	28.84	3820	0.48
Ocotea rubra Mez	623	23.47	2521	5.42
Khaya grandifoliola C.DC.	646	33.02	2968	0.95
Khaya senegalensis A. Juss.	792	33.98	3101	0.33
Coula edulis Baill.	1048	27.6	2674	2.1
Tarrietia javanica Bl.	780	20.33	2198	9.15
Entandrophragma cylindricum Sprague	734	30.6	2592	1.12
Afzelia pachyloba Harms	742	20.56	2048	8.24
Swietenia macrophylla King	571	20.99	1991	9.22
Aucoumea klaineana Pierre	399	32.17	2275	1.81
Humbertia madagascariensis Lamk	1277	23.36	3171	3.48
Faucherea thouvenotii H. Lec.	1061	20.18	2512	6.05
Ceiba pentandra Gaertn.	299	29.16	2396	2.57
Letestua durissima H. Lec.	1046	19.56	2770	3.87
Monopetalanthus heitzii Pellegr.	466	23.98	2344	5.57
Commiphora sp.	390	16.52	1269	9.77
Dalbergia sp.	916	14.29	2224	9.79
Hymenolobium sp.	600	20.58	2402	7.86
Pseudopiptadenia suaveolens Brenan	875	20.8	1989	6.53
Parkia nitida Miq.	232	26.86	1440	5.75
Bagassa guianensis Aubl.	1076	20.68	2059	6.82
Discoglypremna caloneura Prain	406	34.27	1506	1.38
Brachylaena ramiflora Humbert	866	21.85	2258	4.71
Simarouba amara Aubl.	455	21.26	1654	9.37
Gossweilerodendron balsamiferum Harms	460	35.26	1712	1.08
Manilkara mabokeensis Aubrev.	944	23.89	1788	3.25
Shorea-rubro squamata Dyer	569	23.9	1604	6.75
Autranella congolensis A. Chev.	956	38.97	3380	0.35
Entandrophragma angolense C. DC.	473	22.79	1612	7.67
Distemonanthus benthamianus Baill.	779	19.77	2088	8.75
Terminalia superba Engl. & Diels	583	21.89	2004	9.32
Nesogordonia papaverifera R.Cap.	768	27.96	2097	2.37
Albizia ferruginea Benth.	646	24.71	2221	4.32
Gymnostemon zaizou. Aubrev. & Pellegr.	380	30.15	2130	1.83
Anthonotha fragrans Exell & Hillcoat	777	24.87	1926	4.2
Piptadeniastrum africanum Brenan	975	22.41	3226	3.68
Guibourtia ehie J. Leon.	783	26.36	2156	4.05
Manilkara huberi Standl.	1096	35.11	2692	0.77
Pometia pinnata Forst.	713	25.5	1835	6.23
Glycydendron amazonicum Ducke	627	20.41	2292	7.91
Cunonia austrocaledonica Brong. Gris.	621	31.05	3930	0.59
Nothofagus aequilateralis Steen.	1100	37.76	3028	0.18
Schefflera gabriellae Baill.	570	28.16	1872	1.42
Gymnostoma nodiflorum Johnst.	1189	33	3013	1.26
Dysoxylum sp.	977	23.85	2106	4.49
Calophyllum caledonicum Vieill.	789	19.82	2312	8.66
Gyrocarpus americanus Jacq.	206	38.39	1982	0.6
Pyriluma sphaerocarpum Aubrev.	793	30.83	2318	1.23
Cedrela odorata L.	512	30.45	2070	3
Moronobea coccinea Aubl.	953	21.67	1781	4.92
Goupia glabra Aubl.	885	45.61	2525	0.22
Manilkara huberi Standl.	1187	22.6	2917	2.78
Micropholis venulosa Pierre	665	22.51	3113	7.12
Cedrelinga catenaeformis Ducke	490	22.5	1626	7.31
Vouacapoua americana Aubl.	882	23.18	1986	6.88
Tarrietia Densiflora Aubrev & Normand	603	29.76	2326	1.62

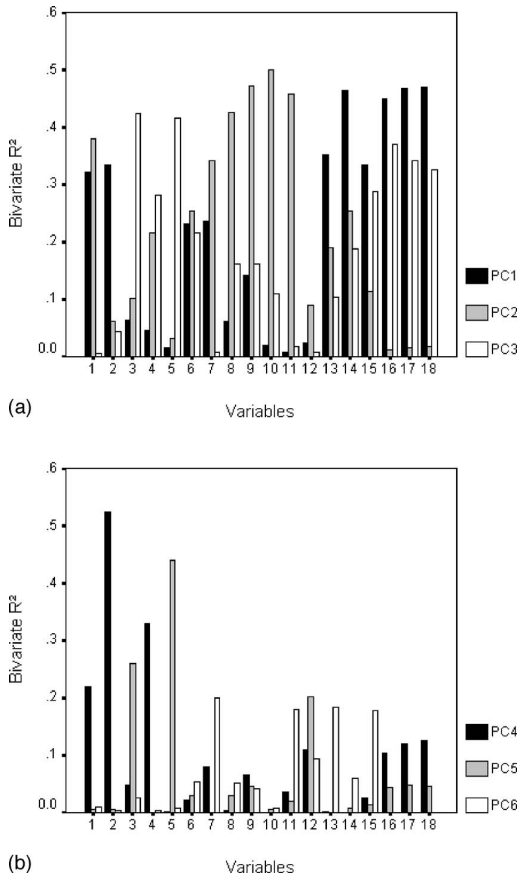


FIG. 8. Bivariate determination coefficient between original variables and principal components: (a) for PC1, PC2 and PC3; (b) for PC4, PC5 and PC6.

entiated by the xylophone maker in the acoustic classification.

b. Multivariate linear regression analysis. The second step of the analysis was to build a robust linear model to take into account the most significant predictors. The robustness of the model assumes that no multicollinearity among the variables exists (Dillon and Goldstein, 1984). The stepwise selection method was thus used to perform multivariate

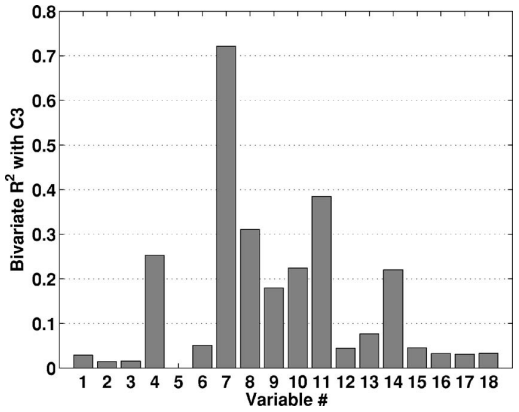


FIG. 9. Bivariate coefficients of determination between characteristic parameters and classification C3 (N=59).

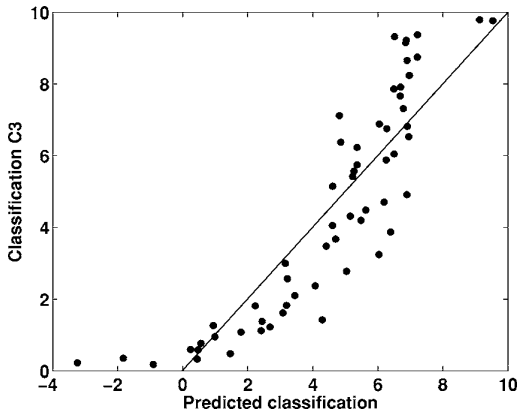


FIG. 10. Predicted vs observed C3 classification (linear predictors α_1 and SB, R^2=0.77, N=59).

analysis. This method enters variables into the model one by one and tests all the variables in the model for removal at each step. Stepwise selection is designed for the case of correlations among the variables. Other automatic selection procedures exist (forward selection and backward elimination, for example), and the models obtained by these methods may differ, especially when independent variables are highly intercorrelated. Because of the high correlation between variables, several regression models almost equally explain classification C3. However, stepwise selection was used to build one of the most significant models with noncorrelated variables relating to different physical phenomena.

The final linear model obtained by stepwise variable selection included the two predictors, α_1 and SB. The predicted classification is given by:

$$\hat{C}3_{\text{Linear}} = -3.82 \times 10^{-1}\alpha_1 - 1.32 \times 10^{-3}SB + 17.52. \tag{20}$$

The multiple coefficient of determination was highly significant (R^2=0.776 and Adjusted R^2=0.768, Fig. 10) and each regression coefficient was statistically different from zero (significance level: 1%). The predictor α_1 was predominant in the model with a partial coefficient value of $R_{\alpha 1}$=−0.84 (R_{SB}=−0.44). The negative sign of $R_{\alpha 1}$ showed that samples with high damping coefficients were associated with a poor acoustic quality.

Partial least squares regression showed that the damping coefficient α_1 was predominant in the model (Brancheriau *et al.*, 2006b). However, the physical significance of the partial least squares model was difficult to explain because the original variables were grouped in latent variables. The stepwise procedure was thus used to better understand the regression results.

The multivariate analysis differed from the bivariate analysis by the replacement of SF by SB, because the selected set of predictors was formed by noncorrelated variables. SB was thus selected because of the low correlation between α_1 and SB with a coefficient value of $R_{\alpha 1/\text{SB}}$=0.29 instead of SF with a value of $R_{\alpha 1/\text{SF}}$=−0.60.

Principal components regression (PCR) was another way to deal with the problem of strong correlations among the variables. Instead of modeling the classification with the variables, the classification was modeled on the principal component scores of the measured variables (which are orthogonal and therefore not correlated). The PCR final model was highly significant with a multiple R^2 value of 0.741 and Adjusted R^2 value of 0.721. Four principal components were selected and the resulting scatter plot was similar to the one in Fig. 10. Comparing the two multivariate models, we found the PCR model to be less relevant than the stepwise one. The R^2 of the PCR model was indeed lower than the R^2 of the stepwise model. Furthermore, the PCR model included four components while only two independent variables were included in the stepwise model. The difference between these two models was explained by the fact that the whole information contained in the characteristic parameters (Table I) was not needed to explain the perceptual classification. The PCR procedure found components that capture the greatest amount of variance in the predictor variables, but did not build components that both capture variance and achieve correlation with the dependent variable.

c. Multivariate nonlinear regression analysis. The configuration of points associated with the linear model (C3, α_1 and SB) in Fig. 10 indicated a nonlinear relationship. This was particularly true for samples of poor acoustic quality (negative values of the standardized predicted classification). As a final step of the analysis, we built a nonlinear model of the behavioral response. In particular, we transformed the values predicted by the linear model $\hat{C}3_{\text{Linear}}$ using a sigmoidal transform. Such transform was consistent with the relationship between C3 and $\hat{C}3_{\text{Linear}}$ (see Fig. 10). The fitting coefficients were extracted via the Levenberg-Marquardt optimization procedure by minimizing the residual sum of squares (dependent variable C3 and independent variable $\hat{C}3_{\text{Linear}}$: predicted classification with the linear modeling). The final equation is written as follows:

$$\hat{C}3_{\text{sigmoid}} = \frac{10}{1 + e^{-\frac{\hat{C}3_{\text{Linear}} - 5}{1.64}}} \tag{21}$$

with $\hat{C}3_{\text{Linear}}$ defined by Eq. (20). The multiple coefficient of determination was highly significant (R^2=0.82) and each nonlinear regression coefficient was statistically different from zero (significance level: 1%). The nonlinear model provided a better fit than the linear model; moreover no apparent systematic feature appeared, indicating that residuals were randomly distributed (Fig. 11).

VII. DISCUSSION

In this section, we discuss the main results presented above, attempting to better understand the sound descriptors' influence on the xylophone maker classification. Further on, we discuss the influence of the pitch and the relationship between the wood anatomy and the produced sounds.

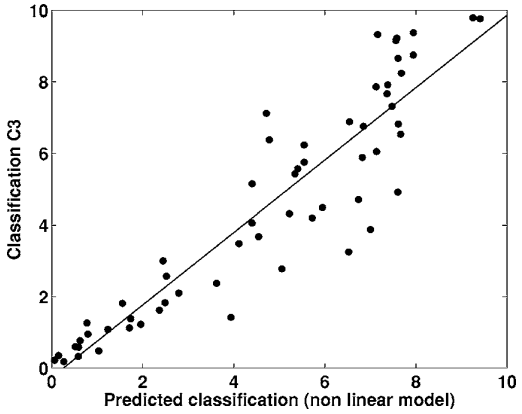


FIG. 11. Predicted vs observed C3 classification (nonlinear predictors α_1 and SB, R^2=0.82, N=59).

A. On the reliability of the xylophone maker

As we pointed out in the introduction, this paper does not aim to give categorical clues for choosing interesting species of wood for xylophone manufacturing. Nevertheless, note that these first conclusions probably accurately reflect what xylophone makers look for. Although we tested our methodology with only one renowned xylophone maker, the results show that:

- In accordance with the xylophone maker community, our maker chose *Dalbergia* sp. as the best species. Moreover, this choice was confirmed on both tuned and original sound classifications.
- The comparison of classifications C1 and C2 showed no significant differences according to the Wilcoxon test.

These observations confirm the good reliability of our xylophone maker and the accuracy of the results, which were further informally confirmed by both instrument makers and musicians.

B. Relation between descriptors and wood classification

The classification by the xylophone maker is correlated with several descriptors. Those that play an important role are three descriptors related to the time course of the sound (α_1, α_2 and SF) and two descriptors related to the spectral content of the sound (SCG and SB). Note that the physical descriptors linked with the wood properties do not explain by themselves the classification of the instrument maker, even though E_ℓ/ρ seems to be the most pertinent one. The relatively low importance of the specific modulus regarding classification C3 could be explained by its high correlation with the fundamental frequency (R^2=0.91) and its low correlation with the temporal damping coefficient α_1 (R^2=0.26). Most of the descriptors are correlated; these correlations are coherent with the physics and are shown in a qualitative way in Fig. 7. Both coefficients of the polynomial decomposition of $\alpha(\omega)$ are strongly correlated. So are the coefficients of the polynomial decomposition of $I(n)$. This finding points out the relative consistency in the behavior of the damping and

the inharmonicity laws with respect to the species. Parameters α_1 and α_2 are also correlated, showing the monotonic behavior of the damping with respect to the frequency: the higher the frequency, the higher the damping. As a consequence, both α_1 and α_2 are correlated with the spectral flux, SF, since these descriptors are the only ones that relate to the time course of the sound.

Both global spectral descriptors, SCG and SB, are also correlated, showing that their increase is strongly related to the adjunction of high frequency energy. These descriptors are in addition correlated with the ratio $A_{2/1}$ and with the physical descriptors ρ and E_ℓ/ρ. This correlation can be explained by the way the energy is distributed through the excited modes. Actually, assuming that the bars are impacted identically (good reproducibility of the impact in the experimental setup), the initial energy injected depends on the impedance of each bar. Since the bars were impacted in the transversal direction, one can assume that the transversal Young modulus of elasticity together with the mass density are the main parameters in the difference of amplitudes of modes 1 and 2.

The multivariate linear regression analysis highlighted two main descriptors: α_1 and SB. These descriptors are noncorrelated and give rise to a linear predictor of the classification $\hat{C}3_{\text{Linear}}$ that explains 77% of the variance. This model is of great importance in the choice of species. Actually, it emphasizes the fact that the xylophone maker looks for a highly resonant sound (the coefficient of α_1 is negative) containing a few spectral components (the coefficient of SB is also negative). Such a search for a crystal-clear sound could explain the general choice of *Dalbergia* sp., which is the most resonant species and the most common in xylophone bars. Indeed, the predominance of α_1 agrees with the first rank of *Dalbergia* sp., for which $\alpha_1 = 14.28\ \text{s}^{-1}$ is the smallest in the data bank ($14.28\ \text{s}^{-1} < \alpha_1 < 45.61\ \text{s}^{-1}$) and SB $= 2224$ Hz is medium range in the data bank (1268 Hz $<$ SB $<$ 3930 Hz). Holz (1996) showed that the damping factor value α_1 should be lower than about $30\ \text{s}^{-1}$ for a fundamental frequency value of 1000 Hz, which corresponds to the mean value of the study. The average value of α_1 is indeed $26.13\ \text{s}^{-1}$ with a standard deviation of $6.18\ \text{s}^{-1}$. Actually, xylophone makers use a specific way of carving the bar by removing substance in the middle (Fletcher and Rossing, 1998). This operation tends to minimize the importance of partial 2, decreasing both the SCG and the SB. The importance of α_1 in the model is in line with several studies showing that the damping is a pertinent clue in the perception of impacted materials (Klatzky *et al.*, 2000; Wildes and Richards, 1988). Concerning parameter SB, the spectral distribution of energy is also an important clue, especially for categorization purposes.

The linear classification prediction has been improved by taking into account nonlinear phenomena. The nonlinear model then explains 82% of the variance. The nonlinear relationship between the perceptual classification and predictors (α_1 and SB) was explained by the instrument maker's strategy during the evaluation of each sample. The xylophone maker proceeded by first identifying the best samples and then the worst samples. This first step gave him the upper and lower bounds of the classification. The final step was to sort the samples of medium quality and place them between the bounds. One could deduce that three groups of acoustic quality (good, poor, and medium quality) were formed before the classification and that inside these groups the perceptual distance between each sample was different. The sigmoid shape indicated that the perceptual distance was shorter for good and poor quality groups than for medium quality groups. As a consequence, the nonlinear model is probably linked with the way the maker proceeded and cannot be interpreted as an intrinsic model for wood classification. Another explanation for the nonlinear relationship can also be found in the nonlinear transform relating physical and perceptual dimensions.

Note finally that there was no correlation between the classification and the wood density. However it is known that the wood density is of great importance for instrument makers. Holz (1996) suggested that the "ideal" xylophone wood bars would have density values between 800 and 950 kg/m^3. This phenomenon is due to the way we designed our experimental protocol, focusing on the sound itself and minimizing multi-sensorial effects (avoiding the access to visual and tactile information). Actually, in a situation where the instrument maker has access to the wood, bars with weak density are rejected for manufacturing and robustness purposes, irrespective of their sound quality.

C. Influence of the fundamental frequency (pitch) on the classification

As discussed previously, timbre is a key feature for appreciating sound quality and it makes it possible to distinguish tones with equal pitch, loudness, and duration (ANSI, 1973). Since this study aims at better understanding which timbre descriptor is of interest for wood classification, one expected differences in the classification of the tuned and the original sound data banks. The difference between classifications C2 (various pitches) and C3 (same pitches) shows a clear linear tendency; it is represented in Fig. 6 as a function of the original fundamental frequency of the bars. The difference is negative (respectively positive) for sounds whose fundamental frequencies are lower (respectively higher) than the mean frequency. The Pearson coefficient associated with the linear relationship between the arithmetic difference of the classification and the fundamental frequency leads to the important observation that *a wooden bar with a low fundamental frequency tends to be upgraded while a wooden bar with a high fundamental frequency tends to be downgraded.* This finding agrees with our linear prediction model, which predicts weakly damped sounds would be better classified than highly damped ones. Actually, sounds with low (respectively high) fundamental frequencies were transposed toward high (respectively low) frequencies during the tuning process, implying α_1 increase (respectively decrease), since the damping is proportional to the square of the frequency (cf. Sec. III C). As an important conclusion, one may say that the instrument maker cannot judge the wood itself independently of the bar dimensions, since the classification is influenced by the pitch changes, favoring wood samples generating low fundamental frequency sounds.

Once again, note the good reliability of our instrument maker, who did not change the classification of sounds whose fundamental frequency was close to the mean fundamental frequency of the data bank (i.e., sounds with nearly unchanged pitch). Actually, the linear regression line passes close to 0 at the mean frequency 1002 Hz. Moreover, the *Dalbergia* sp. was kept at the first position after the tuning process, suggesting that no dramatic sound transformations had been made. In fact, this sample was transposed upwards by 58 Hz, changing α_1 from 13.6 s^{-1} to 14.28 s^{-1}, which still was the smallest value of the tuned data bank.

D. Relationship between wood anatomy and perceived musical quality

The damping α_1 of the first vibrational mode was an important descriptor explaining the xylophone maker classification. Equation (11) shows that this descriptor is related to the quality factor Q, and consequently to the internal friction coefficient $\tan \phi$ (inverse of the quality factor Q), which depends on the anatomical structure of the wood. An anatomical description of the best classified species has been discussed in a companion article (Brancheriau *et al.*, 2006b). We briefly summarize the main conclusions and refer the reader to the article for more information. A draft anatomical portrait of a good acoustic wood could be drawn up on the basis of our analysis of wood structures in the seven acoustically best and seven poorest woods. This portrait should include a compulsory characteristic, an important characteristic, and two or three others of lesser importance. The key trait is the axial parenchyma. It should be paratracheal, and not very abundant if possible. If abundant (thus highly confluent), the bands should not be numerous. Apotra-cheal parenchyma can be present, but only in the form of well-spaced bands (e.g., narrow marginal bands). The rays (horizontal parenchyma) are another important feature. They should be short, structurally homogeneous but not very numerous. The other characteristics are not essential, but they may enhance the acoustic quality. These include:

- Small numbers of vessels (thus large);
- A storied structure;
- Fibers with a wide lumen (or a high flexibility coefficient, which is the ratio between the lumen width and the fiber width; it is directly linked with the thickness of the fiber).

These anatomical descriptions give clues for better choosing wood species to be used in xylophone manufacturing. They undoubtedly are valuable for designing new musical materials from scratch, such as composite materials.

VIII. CONCLUSION

We have proposed a methodology associating analysis-synthesis processes and perceptual classifications to better understand what makes the sound produced by impacted wooden bars attractive for xylophone makers. This methodology, which focused on timbre-related acoustical properties, requires equalization of the pitch of recorded sounds. Statistical analysis of the classifications made by an instrument maker highlighted the importance of two salient descriptors: the damping of the first partial and the spectral bandwidth of the sound, indicating he searched for highly resonant and crystal-clear sounds. Moreover, comparing the classifications of both the original and processed sounds showed how the pitch influences the judgment of the instrument maker. Indeed, sounds with originally low (respectively high) fundamental frequency were better (lesser) classified before the tuning process than after. This result points to the preponderance of the damping and reinforces the importance of the pitch manipulation to better dissociate the influence of the wood species from that of the bar geometry. Finally, the results revealed some of the manufacturers' strategies and pointed out important mechanical and anatomical characteristics of woods used in xylophone manufacturing. From a perceptual point of view, the internal friction seems to be the most important characteristic of the wood species. Nevertheless, even though no correlation has been evidenced between the classification and the wood density, it is well known that this parameter is of great importance for instrument makers as evidence of robustness. As mentioned in the introduction, this work was the first step towards determining relations linking sounds and wood materials. Future works will aim at confirming the results described in this paper by taking into account classifications made by other xylophone makers in the statistical analysis. We plan to use this methodology on a new set of wood species having mechanical and anatomical characteristics similar to those well classified in the current test. This should point out unused wood species of interest to musical instrument manufacturers and will give clues for designing new musical synthetic materials.

ACKNOWLEDGMENTS

The authors thank Robert Hébrard, the xylophone maker who performed the acoustic classification of the wood species. They are also grateful to Pierre Détienne for useful advice and expertise in wood anatomy. They also thank Bloen Metzger and Dominique Peyroche d'Arnaud for their active participation in the experimental design and the acoustical analysis, and Jérémy Marozeau who provided the graphical interface for the listening test. We would also thank the reviewers for useful suggestions.

American National Standards Institute (**1973**). *American National Standard Psychoacoustical Terminology* (American National Standards Institute, NY).

Adrien, J. M. (**1991**). *The Missing Link: Modal Synthesis* (MIT Press, Cambridge, MA), Chap. 8, pp. 269–297.

Avanzini, F., and Rocchesso, D. (**2001**). "Controlling material properties in physical models of sounding objects," in *Proceedings of the International Computer Music Conference 2001*, 17–22 September 2001, Hawana, pp. 91–94.

Beauchamps, J. W. (**1982**). "Synthesis by spectral amplitude and "brightness" matching of analyzed musical instrument tones," J. Audio Eng. Soc. **30**(6), 396–406.

Blay, M., Bourgain , and Samson (**1971**). "Application des techniques électroacoustiques à la détermination du module d'élasticité par un procédé nondestructif (Application of electroacoustic techniques to determine the elasticity modulus by nondestructive procedure)," Technical Review to Advance Techniques in Acoustical, Electrical and Mechanical Measurement **4**, 3–19.

Bork, I. (**1995**). "Practical tuning of xylophone bars and resonators," Appl. Acoust. **46**, 103–127.

Brancheriau, L., and Baillères, H. (**2002**). "Natural vibration analysis of

clear wooden beams: A theoretical review," Wood Sci. Technol. **36**, 347–365.
Brancheriau, L., Baillères, H., Détienne, P., Gril, J., and Kronland-Martinet, R. (**2006a**). "Key signal and wood anatomy parameters related to the acoustic quality of wood for xylophone-type percussion instruments," J. Wood Sci. **52**(3), 270–274.
Brancheriau, L., Baillères, H., Détienne, P., Kronland-Martinet, R., and Metzger, B. (**2006b**). "Classifying xylophone bar materials by perceptual, signal processing and wood anatomy analysis," Ann. Forest Sci. **62**, 1–9.
Bucur, V. (**1995**). *Acoustics of Wood* (CRC Press, Berlin).
Caclin, A., McAdams, S., Smith, B. K., and Winsberg, S. (**2005**). "Acoustic correlates of timbre space dimensions: A confirmatory study using synthetic tones," J. Acoust. Soc. Am. **118**(1), 471–482.
Chaigne, A., and Doutaut, V. (**1997**). "Numerical simulations of xylophones. I. Time-domain modeling of the vibrating bars," J. Acoust. Soc. Am. **101**(1), 539–557.
Dillon, W. R., and Goldstein, M. (**1984**). *Multivariate Analysis—Methods and Applications* (Wiley, New York).
Doutaut, V., Matignon, D., and Chaigne, A. (**1998**). "Numerical simulations of xylophones. II. Time-domain modeling of the resonator and of the radiated sound pressure," J. Acoust. Soc. Am. **104**(3), 1633–1647.
Fletcher, N. H., and Rossing, T. D. (**1998**). *The Physics of Musical Instruments*, 2nd ed. (Springer-Verlag, Berlin).
Giordano, B. L., and McAdams, S. (**2006**). "Material identification of real impact sounds: Effects of size variation in steel, wood, and Plexiglass plates," J. Acoust. Soc. Am. **119**(2), 1171–1181.
Graff, K. F. (**1975**). *Wave Motion in Elastic Solids* (Ohio State University Press), pp. 100–108.
Holz, D. (**1996**). "Acoustically important properties of xylophon-bar materials: Can tropical woods be replaced by European species?" Acust. Acta Acust. 82(6), 878–884.
Klatzky, R. L., Pai, D. K., and Krotkov, E. P. (**2000**). "Perception of material from contact sounds," Presence: Teleoperators and Virtual Environments **9**(4), 399–410.
Lutfi, R. A., and Oh, E. L. (**1997**). "Auditory discrimination of material changes in a struck-clamped bar," J. Acoust. Soc. Am. **102**(6), 3647–3656.
Marozeau, J., de Cheveigné, A., McAdams, S., and Winsberg, S. (**2003**). "The dependency of timbre on fundamental frequency," J. Acoust. Soc. Am. **114**, 2946–2957.
Matsunaga, M., and Minato, K. (**1998**). "Physical and mechanical properties required for violin bow materials II. Comparison of the processing properties and durability between pernambuco and substitutable wood species," J. Wood Sci. **44**(2), 142–146.
Matsunaga, M., Minato, K., and Nakatsubo, F. (**1999**). "Vibrational property changes of spruce wood by impregnating with water-soluble extractives of pernambuco (*Guilandina echinata Spreng*.)," J. Wood Sci. **45**(6), 470–474.
Matsunaga, M., Sugiyama, M., Minato, K., and Norimoto, M. (**1996**). "Physical and mechanical properties required for violin bow materials," Holzforschung **50**(6), 511–517.
McAdams, S., Chaigne, A., and Roussarie, V. (**2004**). "The psychomechanics of simulated sound sources: Material properties of impacted bars," J. Acoust. Soc. Am. **115**(3), 1306–1320.
McAdams, S., Winsberg, S., Donnadieu, S., Soete, G. D., and Krimphoff, J. (**1995**). "Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes," Psychol. Res. **58**, 177–192.
Obataya, E., Umewaza, T., Nakatsubo, F., and Norimoto, M. (**1999**). "The effects of water soluble extractives on the acoustic properties of reed (*Arundo donax* L.)," Holzforschung **53**(1), 63–67.
Ono, T., and Norimoto, M. (**1983**). "Study on Young's modulus and internal friction of wood in relation to the evaluation of wood for musical instruments," Jpn. J. Appl. Phys., Part 1 **22**(4), 611–614.
Ono, T., and Norimoto, M. (**1985**). "Anisotropy of Dynamic Young's Modulus and Internal Friction in Wood," Jpn. J. Appl. Phys., Part 1 **24**(8), 960–964.
Steiglitz, K., and McBride, L. E. (**1965**). "A technique for the identification of linear systems," IEEE Trans. Autom. Control **AC-10**, 461–464.
Sugiyama, M., Matsunaga, M., Minato, K., and Norimoto, M. (**1994**). "Physical and mechanical properties of pernambuco (*Guilandina echinata* Spreng.) used for violin bows," Mokuzai Gakkaishi **40**, 905–910.
Valette, C., and Cuesta, C. (**1993**). *Mécanique de la Corde Vibrante (Mechanics of Vibrating String)*, Traité des Nouvelles Technologies, série Mécanique (Hermès, Paris).
Wildes, R. P., and Richards, W. A. (**1988**). *Recovering Material Properties from Sound* (MIT Press, Cambridge, MA), Chap. 25, pp. 356–363.

Controlling the Perceived Material in an Impact Sound Synthesizer

Mitsuko Aramaki, *Member, IEEE*, Mireille Besson, Richard Kronland-Martinet, *Senior Member, IEEE*, and Sølvi Ystad

***Abstract*—In this paper, we focused on the identification of the perceptual properties of impacted materials to provide an intuitive control of an impact sound synthesizer. To investigate such properties, impact sounds from everyday life objects, made of different materials (wood, metal and glass), were recorded and analyzed. These sounds were synthesized using an analysis–synthesis technique and tuned to the same chroma. Sound continua were created to simulate progressive transitions between materials. Sounds from these continua were then used in a categorization experiment to determine sound categories representative of each material (called *typical* sounds). We also examined changes in electrical brain activity (using event related potentials (ERPs) method) associated with the categorization of these typical sounds. Moreover, acoustic analysis was conducted to investigate the relevance of acoustic descriptors known to be relevant for both timbre perception and material identification. Both acoustic and electrophysiological data confirmed the importance of damping and highlighted the relevance of spectral content for material perception. Based on these findings, controls for damping and spectral shaping were tested in synthesis applications. A global control strategy, with a three-layer architecture, was proposed for the synthesizer allowing the user to intuitively navigate in a "material space" and defining impact sounds directly from the material label. A formal perceptual evaluation was finally conducted to validate the proposed control strategy.**

***Index Terms*—Analysis–synthesis, control, event related potentials, impact sounds, mapping, material, sound categorization, timbre.**

I. Introduction

THE current study describes the construction of a synthesizer dedicated to impact sounds that can be piloted using high-level verbal descriptors referring to material categories (i.e., wood, metal and glass). This issue is essential for sound design and virtual reality where sounds coherent with visual scenes are to be constructed. Control strategies for synthesis (also called mapping) is an important issue that has interested the computer music community ever since it became possible to produce music with computers [1]. A large number of interfaces and control strategies have been proposed by several authors [2]–[10]. Most of these interfaces were designed for musical purposes and are generally not adapted to build environmental sounds used in sound design and virtual reality. As opposed to music-oriented interfaces that generally focus on the control of acoustic factors such as pitch, loudness, or rhythmic deviations, a more intuitive control based on verbal descriptors that can be used by non-experts is needed in these new domains. This issue requires knowledge on acoustical properties of sounds and how they are perceived. As a first approach towards the design of such an environmental sound synthesizer, we focus on the class of impact sounds and on the control of the perceived material. In particular, our aim is to develop efficient mapping strategies between words referring to certain material categories (i.e., wood, metal and glass) and signal parameters to allow for an intuitive sound synthesis based on a smaller number of control parameters.

To point out perceptual properties that characterize the categories, a listening test was conducted. Stimuli were created first by recording impact sounds from everyday life objects made of different materials. Then, these recorded sounds were synthesized by analysis–synthesis techniques and tuned to the same chroma. Finally, we created continua from the tuned sounds to simulate progressive transitions between the categories by interpolating signal parameters. The use of sound continua was of interest to closely investigate transitions and limits between material categories. Sounds from these continua were used in a categorization task so as to be classified by participants as Wood, Metal, or Glass. From the percentage of responses, we determined sound categories representative of each material (called sets of *typical* sounds).

Then, we examined the acoustic characteristics that differ across typical Wood, Metal, and Glass sounds. For this purpose, we considered acoustic descriptors known to be relevant both for timbre perception and for material identification. Previous studies on the perception of sound categories have mainly been based on the notion of timbre. Several authors have used dissimilarity ratings to identify timbre spaces in which sounds from different musical instruments can be distinguished [11]–[14]. They found correlations between dimensions of these timbre spaces and acoustic descriptors such as attack time (the way the energy rises at the sound onset), spectral bandwidth (spectrum spread), or spectral centroid (center of gravity of the spectrum). More recently, roughness (distribution of interacting frequency components within the limits of a critical band) was considered

Manuscript received April 27, 2009; revised October 08, 2009; accepted March 18, 2010. Date of publication April 08, 2010; date of current version October 27, 2010. This work was supported by the Human Frontier Science Program under Grant HFSP #RGP0053 to M. Besson and a grant from the French National Research Agency (ANR, JC05-41996, "senSons") to S. Ystad. The work of M. Aramaki was supported by a postdoctoral grant, first from the HFSP and then from the ANR. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Gaël Richard.

M. Aramaki and M. Besson are with CNRS-Institut de Neurosciences Cognitives de la Méditerranée, 13402 Marseille Cedex 20 France, and also with Aix-Marseille-Université, 13284 Marseille Cedex 07 France (e-mail: aramaki@incm.cnrs-mrs.fr; besson@incm.cnrs-mrs.fr).

R. Kronland-Martinet and S. Ystad are with the CNRS-Laboratoire de Mécanique et d'Acoustique, 13402 Marseille Cedex 20 France (e-mail: kronland@lma.cnrs-mrs.fr; ystad@lma.cnrs-mrs.fr).

Digital Object Identifier 10.1109/TASL.2010.2047755

1558-7916/$26.00 © 2010 IEEE

as a relevant dimension of timbre since it is closely linked to the concept of consonance in a musical context [15], [16]. In the case of impact sounds, the perception of material seems mainly to correlate with the frequency-dependent damping of spectral components [17], [18] ([19], [20] in the case of struck bars), due to various loss mechanisms. Interestingly, damping remains a robust acoustic descriptor to identify macro-categories (i.e., between wood–Plexiglas and steel–glass categories) across variations in the size of objects [21]. From an acoustical point of view, a global characterization of the damping can be given by the sound decay measuring the decrease in sound energy as a function of time.

The above-mentioned timbre descriptors, namely attack time, spectral bandwidth, roughness, and normalized sound decay, were considered as potentially relevant signal features for the discrimination between sound categories. An acoustic analysis was conducted on these descriptors to investigate their relevance. At this stage, it is worth mentioning that signal descriptors that are found to be significant in traditional timbre studies may not be directly useful in the case of sound synthesis and control. Some descriptors might not give access to a sufficiently fine control of the perceived material. It might be necessary to act on a combination of descriptors. To more deeply investigate perceptual/cognitive aspects linked to the sound categorization, we exploited electrophysiological measurements for synthesis purposes since they provide complementary information regarding the nature of sound characteristics that contribute to the differentiation of material categories from a perceptual/cognitive point of view. In particular, we examined changes in brain electrical activity [using event related potentials (ERPs)] associated with the perception and categorization of typical sounds (we refer the reader to a related article for more details [22]).

Based on acoustic and electrophysiological results, sound characteristics relevant for an accurate evocation of material were determined and control strategies related to physical and perceptual considerations were proposed. The relevance of these strategies in terms of an intuitive manipulation of parameters was further tested in synthesis applications. High-level control was achieved through a calibration process to determine the range values of the damping parameters specific to each material category. In particular, the use of sound continua in the categorization experiment highlighted transition zones between categories that allowed for continuous control between different materials.

The paper is organized as follows: first the sound categorization experiment with stimuli construction and results is presented. Statistical analyses are further carried out on the set of sounds defined as *typical* to determine the acoustic descriptors that best discriminate sound categories. Then, sound characteristics that are relevant for material perception are obtained from physical considerations, timbre investigations and electrophysiological measurements. Control strategies allowing for an intuitive manipulation of these sound characteristics, based on these findings and on our previous works [23]–[25], are proposed in a three-layer control architecture providing the synthesis of impact sounds directly from the material label. A formal perceptual evaluation of the proposed control strategy is finally presented.

II. Sound Categorization Experiment

A. Participants

Twenty-five participants (13 women and 12 men, 19 to 35 years old, mean age $= 22.5$) were tested in this experiment that lasted for about one hour. They were all right-handed, nonmusicians (no formal musical training), had normal audition and no known neurological disorders. They all gave written consent and were paid to participate in the experiment.

B. Stimuli

We first recorded 15 sounds by impacting everyday life objects made of three different materials (wooden beams, metallic plates, glass bowls) that are five sounds per material. Synthetic versions of these recorded sounds were generated by an analysis–synthesis process and tuned to the same chroma. Then, we created J-step sound continua that simulate progressive transitions between two sounds of different materials by acting on amplitudes and damping parameters. The different stages of the stimuli construction are detailed below.

1) Analysis–Synthesis of Natural Sounds: Recordings of natural sounds were made in an acoustically treated studio of the laboratory using a microphone placed 1 m from the source. The objects from different materials were impacted by hand. We tried to control the impact on the object by using the same drumstick and the same impact force. The impact position on the different objects was chosen so that most modes were excited (near the center of the object for wooden beams and metallic plates; near the rim for glass bowls). Sounds were digitally recorded at 44.1-kHz sampling frequency.

From a physical point of view, the vibrations of an impacted object (under free oscillations) can generally be modeled as a sum of M exponentially damped sinusoids:

$$s(t) = \theta(t) \sum_{m=1}^{M} A_m \sin(\omega_m t + \Phi_m) e^{-\alpha_m t} \tag{1}$$

where $\theta(t)$ is the Heaviside function and the parameters A_m, α_m, ω_m, and Φ_m, the amplitude, damping coefficient, frequency, and phase of the mth component, respectively. Based on the signal model corresponding to (1), we synthesized the recorded sounds at the same sampling frequency. Several different techniques allow precise estimating the signal parameters $\{A_m, \alpha_m, \omega_m\}_{m=1,\ldots,M}$ based on high-resolution analysis such as the Steiglitz–McBride technique [26] or more recently Estimation of Signal Parameters via Rotational Invariance Techniques (ESPRIT), MUltiple SIgnal Classification (MUSIC), Least Squares or Maximum-Likehood techniques [27]–[30] (see also [31], [32]). These latter methods provide an accurate estimation and can be used to conduct spectral analysis. We here used a simplified analysis technique based on discrete Fourier transform (DFT) since we aimed at reproducing the main characteristics of the original sounds in terms of perceived material rather than achieving a perfect resynthesis.

The number of components M to synthesize was estimated from the modulus of the spectral representation of the signal.

Only the most prominent components, which amplitudes were larger than a threshold value fixed at 30 dB below the maximum amplitude of the spectrum, were synthesized. In addition, to keep the broadness of the original spectrum, we made sure that at least the most prominent component in each critical bandwidth was synthesized. Since Wood and Glass sounds had relatively poor spectra (i.e., few components), most of the components were synthesized. By contrast, Metal sounds had rich and broadband spectra. Some components were due to the nonlinear vibrations of the impacted object (favored by a low dissipation for Metal) and could not be reproduced by the signal model that only considers linear vibrations. Thus, the number of components for synthetic Metal sounds were generally inferior to the number of components of the original sound.

The frequency values ω_m were directly inferred from the abscissa of the local maxima corresponding to the prominent components. Since the spectrum was obtained by computing a fast Fourier transform (FFT) over 2^{16} samples, the frequency precision of each component was equal to 0.76 Hz ($= 44\,100/2^{16}$). Each component m was isolated using a gaussian window centered on the frequency ω_m. The frequency bandwidth of the gaussian window was adapted to numerically minimize the smoothing effects and to avoid the overlap of two successive components which causes interference effects. The gaussian window presents the advantage of preserving the exponential damping when convolved with an exponentially damped sine wave. Then, the analytic signal $\hat{s}_m(t)$ of the windowed signal was calculated using the Hilbert transform and the modulus of $\hat{s}_m(t)$ was modeled by an exponentially decaying function

$$|\hat{s}_m(t)| = A_m e^{-\alpha_m t} \tag{2}$$

Thus, by fitting the logarithm of $|\hat{s}_m(t)|$ with a polynomial function of degree 1 at best in a least-squares sense, the amplitude A_m was inferred from the ordinate at the origin while the damping coefficient α_m was inferred from the slope. Finally, the phases Φ_m were set to 0 for all components. This choice is commonly adopted in synthesis processes since it avoids undesirable clicks at sound onset. It is worth noticing that this phase adjustment does not affect the perception of the material because phase relationships between components mainly reflect the position of the microphone relative to the impacted object.

2) Tuning: The pitches of the 15 synthetic sounds (five per material category) differed since they resulted from impacts on various objects. Consequently, sounds were tuned to the same chroma to minimize pitch variations. Tuning is needed to build homogeneous sound continua with respect to pitch (Section II-B4) and to accurately investigate acoustic descriptors (Section III). In particular, the relationships between descriptors will be better interpreted if they are computed on a set of tuned sounds with normalized pitches rather than on a set of sounds with various pitch values.

We first defined the initial pitch of the sounds from informal listening tests: four participants (different from those who participated in the categorization experiment) listened to each sound and were asked to evaluate the pitch by playing the matching note on a piano keyboard. For each sound, the pitch was defined by the note that was most often associated with the sound. Thus, we defined the pitches $G\sharp 3$ (fundamental frequency of 415.30 Hz), $C\sharp 3$ (277.18 Hz), $F\sharp 3$ (369.99 Hz), $C\sharp 3$, and $C\sharp 3$ for the 5 Wood sounds; $A3$ (440.00 Hz), $F\sharp 3$, $D5$ (1174.65 Hz), $E4$ (659.25 Hz), and $E3$ (329.62 Hz) for the 5 Metal sounds, and $C5$ (1046.50 Hz), $E6$ (2637.02 Hz), $C\sharp 6$ (2217.46 Hz), $D5$, and $F5$ (1396.91 Hz) for the 5 Glass sounds. Then, we tuned the sounds to the closest note C with respect to the initial pitch to minimize signal transformations applied on the sounds: Wood sounds were tuned to the pitch C3, Metal sounds to C3 and C4 and Glass sounds to C5 and C6. Therefore, sounds differed by 1, 2, or 3 octaves depending upon the material. Based upon previous results showing high similarity ratings for tone pairs that differed by octaves [33], an effect known as the octave equivalence, we assume that the octave differences between sounds belonging to a same category should have little influence on sound categorization.

In practice, tuned sounds were generated using the previous synthesis technique [(1)]. The amplitudes and phases of components were kept unchanged but the frequencies (noted $\tilde{\omega}_m$ for tuned sounds) and damping coefficients (noted $\tilde{\alpha}_m$) were recalculated as follows. The tuned frequencies $\tilde{\omega}_m$ were obtained by transposing original ones $\{\omega_m\}_{m=1,\ldots,M}$ with a dilation factor η defined from the fundamental frequency values (in Hz), noted F and $\tilde{F}$, of the sound pitches before and after tuning, respectively,

$$\tilde{\omega}_m = \eta \omega_m \quad \text{with} \quad \eta = \frac{\tilde{F}}{F}. \tag{3}$$

The damping coefficient $\tilde{\alpha}_m$ of each tuned component was recalculated by taking into account the frequency-dependency of the damping. For instance, it is known that in case of wooden bars, the damping coefficients increase with frequency following an empirical expression of a parabolic form where parameters depend on the wood species [34]–[36]. To achieve our objectives, we defined a general expression of a damping law $\alpha(\omega)$ chosen as an exponential function

$$\alpha(\omega) = e^{(\alpha_G + \alpha_R \omega)}. \tag{4}$$

The exponential expression presents the advantage of easily fitting various and realistic damping profiles with a reduced number of parameters. $\alpha(\omega)$ is defined by two parameters α_G and α_R characteristic of the intrinsic properties of the material. The parameter α_G reflects global damping and the parameter α_R reflects frequency-relative damping (i.e., difference between high-frequency component damping and low-frequency component damping). Thus, a damping law $\alpha(\omega)$ was estimated on the original sound by fitting the damping coefficients $\{\alpha_m\}_{m=1,\ldots,M}$ with the (4) at best in a least-squares sense. Then, the damping coefficient $\tilde{\alpha}_m$ of the mth tuned component was recalculated according to this damping law (see also [37])

$$\tilde{\alpha}_m = \alpha(\tilde{\omega}_m). \tag{5}$$

3) Gain Adjustment: Sounds were equalized by gain adjustments to avoid the influence of loudness in the categorization judgments. The gain adjustments were determined on the basis of a pretest with four participants (different from those who participated in the categorization experiment). They were asked to balance the loudness level of the tuned sounds. These

tuned sounds were previously normalized by a gain of reference $\Gamma_0 = 1.5 \times A$ with A corresponding to the largest value of the maxima of the signal modulus among the 15 tuned sounds. The coefficient 1.5 is a safety coefficient commonly used in gain adjustment tests to avoid the saturation of the signals after the adjustment. The gain values Γ to be applied on the 5 Wood sounds were equal to [70, 20, 30, 15, 30], on the 5 Metal sounds were equal to [3.5, 1.1, 1, 1.5, 1.3] and on the five Glass sounds were equal to [35, 15, 15, 30, 10].

Finally, the four participants were asked to evaluate the final sounds in terms of perceived material. Results showed that sounds were categorized in the same material category as the original sounds by all participants thereby showing that the main characteristics of the material were preserved.

4) Sound Continua: To closely investigate transitions between material categories, we created 15 J-step sound continua noted Ω_i with five continua for each material transition. The five Wood-Metal continua were indexed from Ω_1 to Ω_5, the five Wood–Glass continua from Ω_6 to Ω_{10} and finally, the five Glass-Metal continua from Ω_{11} to Ω_{15}. Each continuum was composed of 22 hybrid sounds ($J = 22$) that were obtained by mixing the spectra and by interpolating the damping laws of the two extreme sounds. We chose to mix spectra to fix the values of the frequency components which allows minimizing pitch variations across sounds within a continuum (it is known that shifting components modifies pitch). We chose to interpolate damping laws to gradually modify the damping that conveys fundamental information on material perception. Thus, the sound H_j(t) at step j of the continuum is expressed by

$$H_j(t) = \gamma_1(j)\frac{\Gamma_1}{\Gamma_0}\sum_{m=1}^{M} A_m \sin(\omega_m t)e^{-\alpha^j(\omega_m)t} + \gamma_2(j)\frac{\Gamma_2}{\Gamma_0}\sum_{n=1}^{N} A_n \sin(\omega_n t)e^{-\alpha^j(\omega_n)t} \quad (6)$$

where $\{A_m, \omega_m\}_{m=1,\ldots,M}$ and $\{A_n, \omega_n\}_{n=1,\ldots,N}$ correspond to the sets of amplitudes and frequencies of the two extreme sounds and j varies from 1 to 22. The gains Γ_1 and Γ_2 correspond to the gains of the extreme sounds defined from the gain adjustment test according to a gain of reference Γ_0 (see Section II-B2). The gains $\gamma_1(j)$ and $\gamma_2(j)$ vary at each step j on a logarithmic scale, according to the dB scale

$$\gamma_1(j) = 1 - \frac{\log(j)}{\log(J)}$$
$$\gamma_2(j) = 1 - \frac{\log(J-j+1)}{\log(J)}. \quad (7)$$

The damping variation along the continua is computed by interpolating the damping parameters α_G and α_R of the damping law [defined in (4)] estimated on the two extreme sounds (located at step $j = 1$ and $j = 22$, respectively), leading to the determination of a hybrid damping law $\alpha^j(\omega)$ that progressively varies at each step j of the continuum (see Fig. 1)

$$\alpha^j(\omega) = e^{(\alpha_G^j + \alpha_R^j \omega)} \quad (8)$$

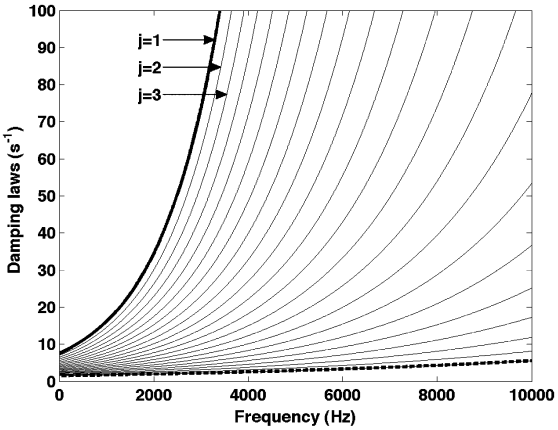


Fig. 1. Damping laws $\alpha^j(\omega)$ for $j = 1, \ldots, 22$ as a function of frequency corresponding to a Wood–Metal continuum. Bold curves correspond to damping laws of Wood (in bold plain) and Metal (in bold dashed) sounds at the extreme positions.

with

$$\alpha_G^j = \left(\alpha_G^{22} - \alpha_G^1\right)\frac{j-1}{J-1} + \alpha_G^1$$
$$\alpha_R^j = \left(\alpha_R^{22} - \alpha_R^1\right)\frac{j-1}{J-1} + \alpha_R^1. \quad (9)$$

The use of an interpolation process on the damping allowed for a better merging between the extreme sounds since the spectral components of the two spectra are damped following the same damping law $\alpha^j(\omega)$. As a consequence, hybrid sounds (in particular, at centered positions of the continua) differed from sounds obtained by only mixing the extreme sounds.

The obtained sounds had different signal lengths (Metal sounds are longer than Wood or Glass sounds). To restrain the lengths to a maximum of 2 seconds, sound amplitudes were smoothly dropped off by multiplying the temporal signal with the half decreasing part of a Hann window.

A total of 330 sounds were created. The whole set of sounds are available at [38]. The averaged sound duration was 861 ms for all sounds and 1053 ms in the Wood–Metal continua, 449 ms in the Wood-Glass continua, and 1081 ms in the Glass–Metal continua.

C. Procedure

The experiment was conducted in a quiet Faradized (electrically shielded) room. Sounds were presented once (i.e., no repetition of the same sound) in random order through one loudspeaker (Tannoy S800) located 1 m in front of the participant. Participants were asked to categorize sounds as Wood, Metal, or Glass, as fast as possible, by pressing one response button out of three on a three-buttons response box[1] (right, middle, and left buttons; one button per material category label). The association between response buttons and material categories was balanced across participants to avoid any bias linked with the

[1]Since participants were not given the option to choose that sounds did not belong to either one of these three categories, results may be biased, but this potential ambiguity would be raised only for intermediate sounds.

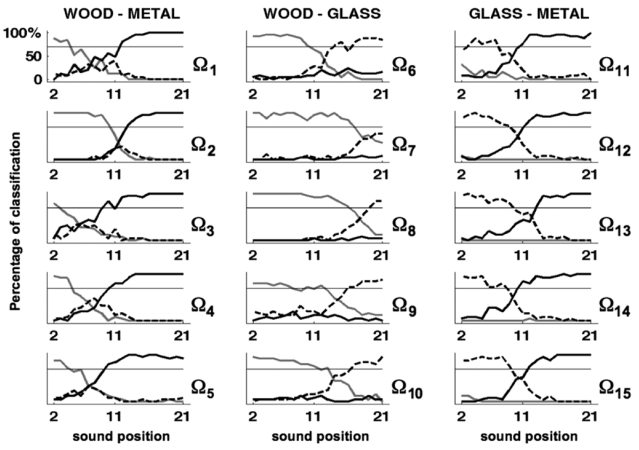


Fig. 2. Percentage of classification as Wood (gray curves), Metal (black curves), or Glass (dashed curves) for each sound as a function of its position j on the continuum for the 15 continua Ω_i. Sounds were considered as typical if they were classified in one category by more than 70% of participants (threshold represented by an horizontal line). No data were collected for extreme sounds ($j = 1$ and $j = 22$) since they were used in the training session.

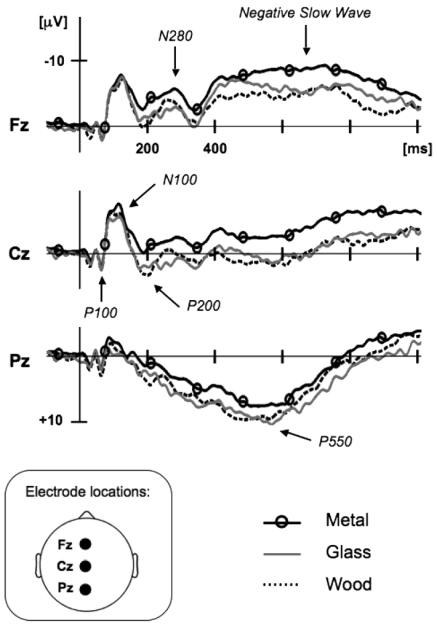


Fig. 3. ERPs to typical sounds of Wood (dotted), Metal (circle marker), and Glass (gray) at electrodes (Fz, Cz, Pz) along the midline of the scalp (see [22] for lateral electrodes). The amplitude (in microvolts) is represented on the ordinate and negativity is up. The time from sound onset is on the abscissa (in milliseconds).

order of the buttons. A row of 4 crosses, i.e., "XXXX," was presented on the screen 2500 ms after sound onset during 1500 ms to give participants time to blink. The next sound was presented after a 500-ms silence. Participants' responses were collected for all sounds except for the extreme sounds since they were used in the training session. The electrical brain activity (Electroencephalogram, EEG) was recorded continuously for each participant during the categorization task.

D. Results

1) Behavioral Data: Percentages of categorization as Wood, Metal, or Glass were obtained for each sound by averaging responses across participants. Fig. 2 allows visualization of these data as a function of sound position along the continua. From these data, we determined a set of *typical* sounds for each material category: sounds were considered as *typical* if they were categorized as Wood, Metal, or Glass by more than 70% of participants (we refer the reader to [39] for more details on the determination of this threshold of percentage value). In addition, sound positions delimiting categories along the continua can be defined from the position ranges of typical sounds. Note that due to the different acoustic characteristics of sounds, the category limits are not located at the same position for all continua (see Fig. 2).

2) Electrophysiological Data: We examined ERPs time-locked to sound onset to analyze the different stages of information processing as they unfold in real time.[2] ERP data were averaged separately for typical sounds of each material category (Fig. 3). We summarize here the main findings (we refer the reader to a related article for more details [22]).

Typical sounds from each category elicited small P100 components, with maximum amplitude around 65-ms post-sound onset, large N100, and P200 components followed by N280 components and Negative Slow Wave (NSW) at fronto-central sites or large P550 components at parietal sites. Statistical analyses revealed no significant differences on the P100 and N100 components as a function of material categories. By contrast, they showed that typical Metal sounds elicited smaller P200 and P550 components, and larger N280 and NSW components than typical sounds from Wood and Glass. From an acoustic point of view, Metal sounds have richer spectra and longer durations (i.e., lower damping) than Wood and Glass sounds. The early differences on the P200 components most likely reflect the processing of spectral complexity (see [42] and [43]) while the later differences on the N280 and on the NSW are likely to reflect differences in sound duration (i.e., differences in damping; see [44] and [45]).

III. Acoustic Analysis

The typical sounds as defined based upon behavioral data (Section II-D1) form a set of sounds representative of each material category. To characterize these typical sounds from an acoustical point of view, we investigated the following descriptors: attack time (AT), spectral bandwidth (SB), roughness (R), and normalized sound decay (α) that are defined below. Then, we examined the relationships between acoustic descriptors and their relevance to discriminate material categories.

A. Definition of Acoustic Descriptors

Attack time is a temporal timbre descriptor which characterizes signal onset. It is defined by the time (in second) necessary

[2]The ERPs elicited by a stimulus (a sound, a light, etc.) are characterized by a succession of positive (P) and negative (N) deflections relative to a baseline (usually measured within the 100 ms or 200 ms that precedes stimulus onset). These deflections (called components) are characterized by their polarity, their latency of maximum amplitude (relative to stimulus onset), their distribution across different electrodes located at standard positions on the scalp and by their functional significance. Typically, the P100, N100, and P200 components reflect the sensory and perceptual stages of information processing, and are obligatory responses to the stimulation [40], [41]. Then, depending on the experimental design and on the task at hand, different late ERP components are elicited (N200, P300, N400, etc.).

for the signal energy to raise from a threshold level to the maximum energy in the temporal envelope (for percussive sound) or to the sustained part (for a sustained sound with no decay part) [46], [47]. Different values have been proposed in the literature for both minimum and maximum thresholds. For our concern, we chose to compute the attack time from 10% to 90% of the maximum amplitude of the temporal envelope as in [48]. This descriptor is known to be relevant to distinguish different classes of instrumental sounds. For instance, sounds from percussive and woodwind instruments have respectively short and long AT.

Spectral bandwidth (in Hz), commonly associated with the spectrum spread, is defined by [49]

$$SB = \frac{1}{2\pi}\sqrt{\frac{\sum_k |\hat{s}(k)|\left(\omega(k) - 2\pi \times SC\right)^2}{\sum_k |\hat{s}(k)|}} \tag{10}$$

where SC is the spectral centroid (in Hz) defined by [50]

$$SC = \frac{1}{2\pi}\frac{\sum_k \omega(k)\,|\hat{s}(k)|}{\sum_k |\hat{s}(k)|} \tag{11}$$

and where ω represents frequency, $\hat{s}$ the Fourier transform of the signal estimated using the FFT algorithm and k the FFT bin index. The FFT was calculated on 2^{16} samples.

Roughness (in asper) is commonly associated with the presence of several frequency components within the limits of a critical band. From a perceptual point of view, roughness is correlated with tonal consonance based on results from experiments on consonance judgments conducted by [51]. From a signal point of view, [52] have shown that roughness and fluctuation strength are proportional to the square of the modulation factor of an amplitude modulated pure tone. We computed roughness based on Vassilakis's model by summing up the partial roughness r_{mn} for all pairs of frequency components contained in the sound [53]

$$r_{mn} = 0.5(A_m A_n)^{0.1} \times \left(\frac{2\min(A_m, A_n)}{A_m + A_n}\right)^{3.11} \times \left(e^{-3.5v|\omega_m - \omega_n|} - e^{-5.75v|\omega_m - \omega_n|}\right) \tag{12}$$

with

$$v = \frac{0.24}{0.0207 \times \min(\omega_m, \omega_n) + 2\pi \times 18.96}$$

and where A_m and A_n are amplitudes and ω_m and ω_n are the frequencies of components m and n, respectively.

Finally, the sound decay D (in s^{-1}) quantifies the amplitude decrease of the whole temporal signal and globally characterizes the damping in the case of impact sounds. In particular, D approximately corresponds to the decay of the spectral component with the longest duration (i.e., generally the lowest frequency one). The sound decay is directly estimated by the slope of the logarithm of the temporal signal envelope. This envelope is given by calculating the analytic signal using the Hilbert transform and by filtering the modulus of this analytic signal using a second-order low-pass Butterworth filter with cutoff frequency of 50 Hz [36]. Since damping is frequency dependent (Section II-B1), sound decay depends on the spectral content of the sound. Consequently, we considered a normalized sound decay denoted α with respect to the spectral localization of the energy and we defined the dimensional descriptor α as the ratio of the sound decay D to the SC value

$$\alpha = \frac{D}{SC}. \tag{13}$$

TABLE I
COEFFICIENTS OF DETERMINATION BETWEEN THE ATTACK TIME AT, THE SPECTRAL BANDWIDTH SB, THE ROUGHNESS R, AND THE NORMALIZED SOUND DECAY α. SINCE THE MATRIX IS SYMMETRIC, ONLY THE UPPER PART IS REPORTED. THE P-VALUES ARE ALSO REPORTED BY *** ($p < .001$) WHEN COEFFICIENTS ARE SIGNIFICANT (WITH BONFERRONI ADJUSTMENT)

	AT	SB	R	α
AT	1	0.05^{***}	0.07^{***}	0
SB	–	1	0.25^{***}	0.02
R	–	–	1	0.23^{***}
α	–	–	–	1

B. Relationships Between Acoustic Descriptors

As a first step, we examined the relationships between the acoustic descriptors estimated on typical sounds. Table I shows the coefficients of determination that are the square of the Bravais–Pearson coefficients between pairs of descriptors. We found highest significant correlation (although not high in terms of absolute value) between the two spectral descriptors SB and R. Lowest correlations were found between AT and the other ones, reflecting the fact that sound onset has little influence on the spectral characteristics and does not depend on the decaying part of the sound (described by α).

Second, a principal component analysis (PCA) was conducted on standardized values of acoustic descriptors (i.e., values centered on the mean value and scaled by the standard deviation value). Results showed that the first two principal components explained about 72% of the total variance (the first component alone explained about 48%). As shown in Fig. 4, the first component was mainly correlated to the spectral descriptors (SB and R) and the second component to the temporal descriptors (AT and α). Thus, PCA revealed that sounds could reliably be represented in a reduced bi-dimensional space which orthogonal axes are mainly correlated to spectral (Component I) and temporal descriptors (Component II), respectively. This result confirmed that spectral and temporal descriptors bring complementary information on the sound characterization from an acoustic point of view.

C. Discrimination Between Material Categories

We examined the relevance of acoustic descriptors to discriminate material categories using a discriminant canonical analysis. This analysis was conducted using Materials (Wood, Metal, and Glass) as groups and standardized values of acoustic descriptors $\{\mathrm{AT}, \mathrm{SB}, \mathrm{R}, \alpha\}$ as independent variables. Since three sound categories were considered, two discriminant functions that allow for the clearest separation between sound categories were computed (the number of discriminant functions is equal to the number of groups minus one). These

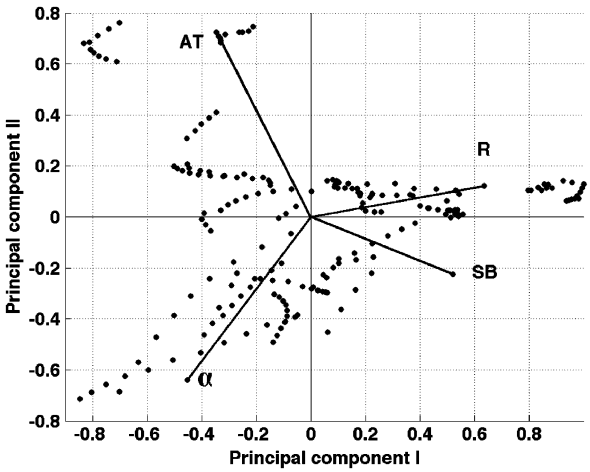


Fig. 4. Biplot visualization: the observations corresponding to typical sounds are represented by dots and the acoustic descriptors by vectors. The contribution of descriptors to each Principal Component (PC) can be quantified by the R^2 statistics given by a regression analysis: Attack time AT ($R^2 = .23$ for PC I and $R^2 = .52$ for PC II), Spectral bandwidth SB ($R^2 = .52$ for PC I and $R^2 = .05$ for PC II), Roughness R ($R^2 = .77$ for PC I and $R^2 = .01$ for PC II) and Normalized sound decay α ($R^2 = .39$ for PC I and $R^2 = .40$ for PC II).

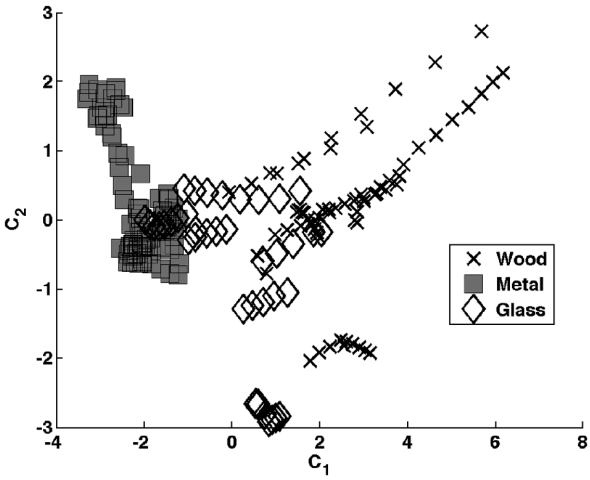


Fig. 5. Scatter plot of the canonical variables allowing the clearest separation between typical Wood (×), Metal (□), and Glass (◇) sound categories.

functions C_1 and C_2 were expressed as a combination of the independent variables

$$C_1 = 1.04\alpha + 0.76SB - 0.58R + 0.47AT$$
$$C_2 = 0.70R - 0.15SB - 0.56AT + 0.38\alpha. \qquad (14)$$

The Wilks's Lambda show that both functions C_1 (Wilks's $\Lambda = .15$; $\chi^2 = 366.65$; $p < .001$) and C_2 (Wilks's $\Lambda = .87$; $\chi^2 = 28.02$; $p < .001$) are significant. The first function C_1 explains 96% of the variance (coefficient of determination $= 0.82$) while the second function C_2 explains the remaining variance (coefficient of determination $= 0.13$). The coefficient associated with each descriptor indicates its relative contribution to the discriminating function. In particular, the first function C_1 is mainly related to α and allows clear distinction particularly between typical Wood and Metal sounds as shown in Fig. 5. This result is in line with previous studies showing that damping is a fundamental cue in the perception of sounds from impacted materials (see the Introduction). The second axis C_2 is mainly related to the spectral descriptor R and allows for a distinction of Glass sounds.

IV. Control Strategy for the Synthesizer

Results from acoustic and electrophysiological data are now discussed in the perspective of designing an intuitive control of the perceived material in an impact sound synthesizer. In particular, we aim at determining relevant sound characteristics for an accurate evocation of different materials and at proposing intuitive control strategies associated with these characteristics. In practice, the synthesis engine and the control strategies were implemented using Max/MSP [54] thereby allowing for the manipulation of parameters in real-time and consequently, providing an easy way to evaluate the proposed controls. The observations and conclusions from these synthesis applications are also reported.

A. Determination of Sound Characteristics

As a starting point, results from acoustic analysis revealed that α (characterizing the damping) was the most relevant descriptor to discriminate material categories, therefore confirming several findings in the literature on the relevance of damping for material identification (see the Introduction). Thus, damping was kept as a relevant sound characteristic to control in the synthesizer.

Furthermore, acoustic analysis showed that in addition to the damping, a second dimension related to spectral characteristics of sounds was significant, in particular for the distinction between Glass and Metal sounds. Interestingly, this result is supported by electrophysiological data that revealed ERP differences between Metal on one side and both Glass and Wood sounds on the other side. Since these differences were interpreted as reflecting processing of sound duration (related to damping) and spectral content, ERP data showed the relevance of both these aspects for material perception. Thus, from a general point of view, it is relevant to assume that material perception seems to be guided by additional cues (other than damping) that are most likely linked to the spectral content of sounds. This assumption is in line with several studies showing the material categorization can be affected by spectral characteristics, in particular, Glass being associated with higher frequencies than Metal sounds [20], [55].

In line with this assumption, synthesis applications confirmed that damping was relevant but was not in some cases sufficient to achieve material categories. For instance, it was not possible to transform a given Wood or Glass sound into a Metal sound by only applying a Metal damping on a Wood or Glass spectrum. The resulting sound did not sound metallic enough. It was therefore necessary to also modify the spectral content, and in particular the spectral distribution, to emphasize the metallic aspect of the sounds (examples in [38]). We found another limitation of damping in the case of Glass synthesis. Indeed, Glass sounds match a wide range of damping values (from highly damped

jar sounds to highly resonant crystal glass sounds) and are most often characterized by a sparse distribution of spectral components (i.e., few distinct components). These observations indicated that the perceptual distinction between Glass and Metal may be due to the typical dissonant aspect of Metal and can be accurately reflected by the roughness that was highlighted as the most relevant descriptor in the acoustic analysis after the damping. Thus, we concluded on the necessity to take into account a control of the spectral shape, in addition to the control of damping, for a more accurate evocation of the perceived material.

Besides, electrophysiological data provided complementary information regarding the temporal dynamics of the brain processes associated with the perception and categorization of typical sounds. First, it is known that P100 and N100 components are influenced by variations in sound onset parameters [56]. The lack of differences on these ERP components is taken to indicate similar brain processes for all typical sounds, showing that the information of the perceived material does not lie in the sound onset. As a synthesis outcome, it means that a modification of sound onset does not affect the nature of the perceived material and consequently, that AT is not a relevant parameter for the control of the perceived material. Second, it is known that N100 component is also influenced by pitch variations [40]. While octave differences were largest between Glass and the other two categories, the lack of differences on N100 component is taken to indicate that pitch may not affect sound categorization. Thus, electrophysiological data support our previous assumption concerning the weak influence linked to the octave differences on sound categorization (Section II-B2).

Based on these considerations, damping and spectral shaping were determined as relevant sound characteristics for material perception. Control strategies associated with these characteristics are proposed and detailed in the following sections.

B. Control of damping

The control of damping was designed by acting on parameters α_G and α_R of the damping law (4). This control gave an accurate manipulation of sound dynamics (time evolution) with a reduced number of control parameters. In particular, the parameter α_G governed the global sound decay (quantified by the descriptor D, Section III) and the parameter α_R allowed controlling the damping differently for high- and low-frequency components. This control made it possible to synthesize a wide variety of realistic sounds. Since from a physical point of view, high frequency components generally are more heavily damped than low frequency ones, we expected both parameters α_G and α_R to be positive in the case of natural sounds.

C. Control of Spectral Shaping

We here propose two control strategies. The first one relied on spectral dilation and was based on physical considerations, in particular on the dispersion phenomena. The second control was based on amplitude and frequency modulations and relied on adding components to the original spectrum. This latter control had a smaller influence on pitch compared with the first control since the original spectrum is not distorted. It also has interesting perceptual consequences. For instance, by creating components within a specific frequency band (i.e., critical band of hearing), this control specifically influenced the perception of roughness that was highlighted as a relevant acoustic descriptor in the acoustic analysis (Section III). These two control strategies are detailed in the following sections.

1) Control by Spectral Dilation: From the analysis of natural sounds, and from physical models describing wave propagation in various media (i.e., various physical materials), two important phenomena can be observed: dispersion and dissipation [32], [57]. Dissipation is due to various loss mechanisms and is directly linked to the damping parameters (α_G and α_R) as described above. Dispersion is linked to the fact that the wave propagation speed varies with respect to frequency. This phenomenon occurs when the phase velocity of a wave is not constant and introduces inharmonicity in the spectrum of the corresponding sound. An example of dispersive medium is the stiff string for which the mth partial is not located at $m\omega_1$ but at $m\omega_1\sqrt{1+\beta m^2}$ where ω_1 is the fundamental frequency and β the coefficient of inharmonicity depending on the physical parameters of the string [58]. We based our first spectral shaping strategy on the spectral dilation defined by

$$\tilde{\omega}_m = W(\breve{\omega}_{\min}, \breve{\omega}_{\max}, \omega)\breve{\omega}_m + (1 - W(\omega_{\min}, \omega_{\max}, \omega))\,\omega_m \tag{15}$$

where W is a window function (defined later in the text) and

$$\breve{\omega}_m = S_G \omega_m \sqrt{1 + S_R \left(\frac{\omega_m}{\omega_1}\right)^2} \tag{16}$$

with ω_m and $\tilde{\omega}_m$ that correspond to the frequency of the initial and shifted component of rank m, respectively. Equation (16) is a generalization of the inharmonicity law previously defined for stiff strings so that the expression is not limited to harmonic sounds but can be applied to any set of frequencies. S_G and S_R are defined as the global and relative shaping parameters, respectively. Ranges of S_G and S_R are constrained so that $\breve{\omega}_m$ are real-valued and $\breve{\omega}_m > \omega_1$ for all $m = 1, \ldots, M$ with M the number of components. Thus, S_R should be lower bounded

$$\min S_R = -\frac{1}{M^2} \tag{17}$$

and S_G should satisfy

$$S_G \frac{\omega_m}{\omega_1} \sqrt{1 + S_R \left(\frac{\omega_m}{\omega_1}\right)^2} > 1 \quad \text{for all } m = 1, \ldots, M. \tag{18}$$

A window function $W(\omega_{\min}, \omega_{\max}, \omega)$ provided a local control of spectral shaping within a given frequency range $[\omega_{\min}; \omega_{\max}]$. In particular, it was of interest to keep the first components unchanged during spectral control to reduce pitch variations. For instance, a window function $W(\omega_3, F_s/2, \omega)$ where F_s is the sampling frequency can be applied to only act on frequencies higher than ω_2. In practice, we chose a Tukey (tapered cosine) window defined between $\omega_{\min}$ and $\omega_{\max}$. The window is parameterized by a ratio ρ (between 0 and 1) allowing the user to choose intermediate profiles from rectangular ($\rho = 0$) to Hann ($\rho = 1$) windows. Consequently, the user is able to act on the weight of the local control.

From an acoustic point of view, the control acts on the spectral descriptors SB and R in a global way. For example, a decrease of the S_G value leads to a decrease of the SB value and at the same time to an increase of the R value.

2) Control by Amplitude and Frequency Modulations: Amplitude modulation creates two components on both sides of the original one and the modulated output waveform is expressed by

$$\begin{aligned} d_m^{\mathrm{AM}}(t) &= A_m\left(1 + I\cos(\omega_n t)\right)\cos(\omega_m t) \\ &= A_m\cos(\omega_m t) + \frac{A_m I}{2}\cos\left((\omega_m + \omega_n)t\right) \\ &\quad + \frac{A_m I}{2}\cos\left((\omega_m - \omega_n)t\right) \end{aligned}$$

where $I \in [0,1]$ is the modulation index, ω_n the modulating frequency, A_m the amplitude, and ω_m the frequency of the mth component.

Frequency modulation creates a set of components on both sides of the original one and the modulated output waveform is expressed by

$$\begin{aligned} d_m^{\mathrm{FM}}(t) &= A_m\cos\left(\omega_m t + I\sin(\omega_n t)\right) \\ &= A_m\sum_{k=-\infty}^{\infty} J_k(I)\cos\left((\omega_m + k\omega_n)t\right) \end{aligned}$$

where $k \in N$ and $J_k(I)$ is the Bessel function of order k. The amplitude of these additional components are given by the amplitude of the original partial A_m and the values of $J_k(I)$ for a given modulation index I.

For both amplitude and frequency modulations, synthesis applications showed that applying the same value of the modulating frequency ω_n to all components led to synthetic sounds perceived as too artificial. To avoid this effect, we proposed a definition of the modulating frequency $\omega_{n,m}$ for each spectral component m based on perceptual considerations. Thus, $\omega_{n,m}$ was expressed as a percentage of the critical bandwidth Δf_m associated with each component m [59]

$$\Delta f_m = 25 + 75\left(1 + 1.4 f_m^2\right)^{0.69} \tag{19}$$

where f_m is expressed in kHz. Since Δf_m increases with respect to frequency, components created at high frequencies are more distant (in frequency) on both sides of the central component than components created at low frequencies. This provided an efficient way to control roughness since the addition of components within a critical bandwidth increases the perception of roughness. In particular, it is known that the maximum sensory dissonance corresponds to an interval between spectral components of about 25% of the critical bandwidth [51], [60].

Synthesis applications showed that both spectral shaping controls allowed for morphing particularly between Glass and Metal sounds while keeping the damping unchanged. In this case, the damping coefficients of the modified frequencies were recalculated according to the damping law. Both controls provided a local control since modifications can be applied on each original component independently. The control based on amplitude and frequency modulations allowed subtle spectral modifications compared with the control based on spectral dilation and in particular, led to interesting fine timbre effects such as cracked glass sounds (sound examples can be found in [38]).

D. Control of the Perceived Material

A global control strategy of the perceived material that integrates the previous damping and spectral shaping controls is proposed in this section. This strategy is hierarchically built on three layers: the "Material space" (accessible to the user), the "Damping and Spectral shaping parameters" and the "Signal parameters" (related to the signal model). Note that the mapping strategy does not depend on the synthesis technique. As a consequence, the proposed control can be applied to any sound generation process.[3] Note also that the proposed strategy is not unique and represents one among several other possibilities [24], [25].

Fig. 6 illustrates the mapping between these three layers based on the first spectral shaping control (using S_G and S_R). The Material space is designed as a unit disk of center C with three fixed points corresponding to the three reference sounds (Wood, Metal, and Glass) equally distributed along the external circle. The Glass sound position is arbitrarily considered as the angle's origin ($\theta = 0$) and consequently, the Metal sound is positioned at $\theta = 2\pi/3$ and the Wood sound at $\theta = 4\pi/3$. The three reference sounds were synthesized from the same initial set of harmonic components (fundamental frequency of 500 Hz and 40 components) so that Wood, Metal, and Glass sounds were obtained by only modifying the damping and spectral shaping parameters (values given in Table II and sound positions shown in Fig. 6). These parameters were chosen on the basis of the sound quality of the evoked material. Note that the reference sounds could be replaced by other sounds.

The user navigates in the Material space between Wood, Metal, and Glass sounds by moving a cursor and can synthesize the sound corresponding to any position. When moving along the circumference of the Material space circle, the corresponding sound $S_h(\theta)$ characterized by its angle θ is generated with Damping and Spectral shaping parameters defined by

$$\mathbf{P}_{S_h}(\theta) = T(\theta)\mathbf{P}_G + T\left(\theta - \frac{2\pi}{3}\right)\mathbf{P}_M + T\left(\theta - \frac{4\pi}{3}\right)\mathbf{P}_W \tag{20}$$

where $\mathbf{P}$ represent the parameter vector $\{\alpha_G, \alpha_R, S_G, S_R\}$ of the sound S_h and of the reference sound of Glass (G), Metal (M), and Wood (W). The function $T(\theta)$ was defined so that the interpolation process was exclusively made between two reference sounds at a time

$$T(\theta)\begin{cases} = -\frac{3}{2\pi}\theta + 1, & \text{for } \theta \in \left[0; \frac{2\pi}{3}\right[\\ = 0, & \text{for } \theta \in \left[\frac{2\pi}{3}; \frac{4\pi}{3}\right] \\ = \frac{3}{2\pi}\theta - 2, & \text{for } \theta \in \left]\frac{4\pi}{3}; 2\pi\right[\end{cases}. \tag{21}$$

Inside the circle, a sound $S'_h(r,\theta)$ characterized by its angle θ and its radius r is generated with parameters defined by

$$\mathbf{P}_{S'_h}(r,\theta) = (1-r)\mathbf{P}_C + r\mathbf{P}_{S_h}(\theta) \tag{22}$$

[3]In practice, we implemented an additive synthesis technique (sinusoids plus noise) in the synthesizer previously developed [23] since it was the most natural one according to the signal model. Other techniques could have been considered as well such as frequency modulation (FM) synthesis, subtractive synthesis, etc.

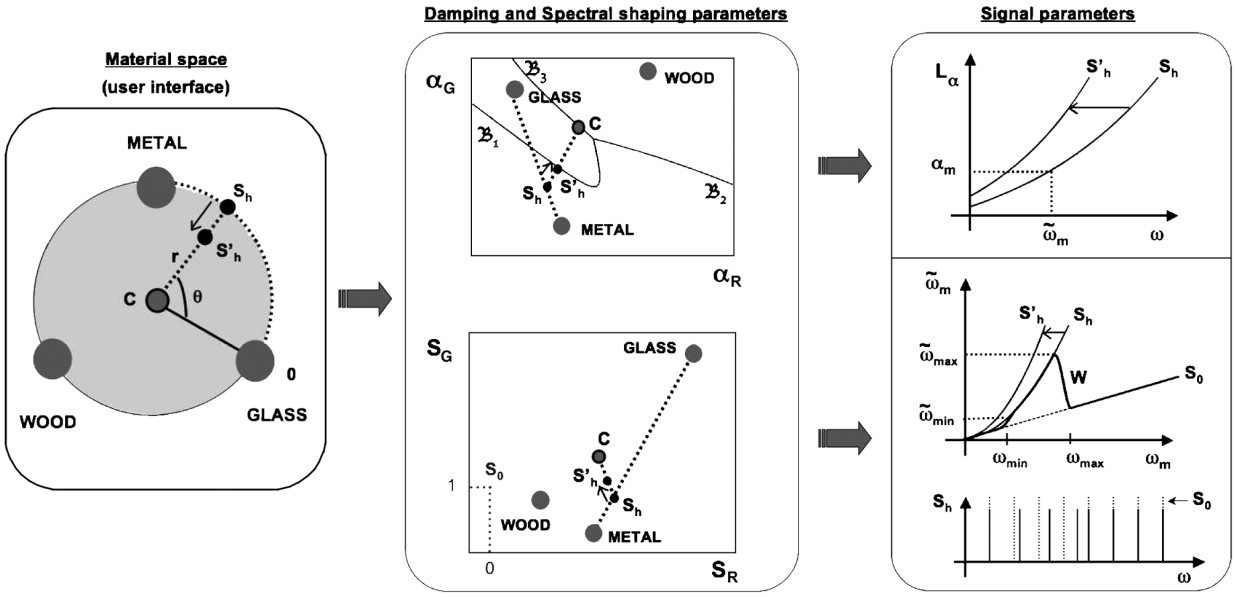


Fig. 6. Control of the perceived material based on three-layer architecture: the "Material space" (user interface), the "Damping and spectral shaping parameters" and "Signal parameters." The navigation in the Material space (e.g., from sound S_h to sound S'_h) involves modifications of both Damping and Spectral shaping parameters. The $\{\alpha_G, \alpha_R\}$ space was calibrated with specific domain for each material category (borders defined in Fig. 7). In this example, we represented the Spectral shaping parameters $\{S_G, S_R\}$ corresponding to the first spectral shaping control. Finally, at signal level, the damping coefficients α_m are computed from (4) with values of $\{\alpha_G, \alpha_R\}$ and the frequencies $\tilde{\omega}_m$ were computed from (15) with values of $\{S_G, S_R\}$. The Metal, Wood, and Glass reference sounds are constructed from the same initial harmonic sound S_0 located at point (0,1) in the $\{S_G, S_R\}$ space. The role of spectral shaping with the window function W is illustrated at the bottom. S_0 is represented in dotted and S_h in bold. The amplitude of the spectrum S_h was arbitrarily reduced for a sake of clarity.

TABLE II
VALUES OF DAMPING (α_G AND α_R) AND SPECTRAL SHAPING (S_G AND S_R) PARAMETERS CORRESPONDING TO THE REFERENCE SOUNDS OF METAL, WOOD, AND GLASS CATEGORY IN THE MATERIAL SPACE

	α_G	α_R ($\times 10^{-4}$)	S_G	S_R
Metal	0.6	2	0.5	0.1
Wood	3	4	0.85	0.05
Glass	2.5	1.5	2.4	0.2

where $\mathbf{P}_C$ represents the parameter vector of the sound C defined by $\{\bar{\alpha}_G, \bar{\alpha}_R, \bar{S}_G, \bar{S}_R\}$ with bar symbol denoting the average of the three values (corresponding to Wood, Metal, and Glass reference sounds) for each parameter and where $\mathbf{P}_{S_h}(\theta)$ is defined in (20). A similar strategy was designed for the mapping based on the second spectral shaping control (amplitude and frequency modulations). In that case, the parameter vector $\mathbf{P}$ corresponded to $\{\alpha_G, \alpha_R, I, \omega_n\}$.

The second layer concerns the controls of Damping and Spectral shaping parameters. For each control, the parameters are represented in two-dimensions to propose an intuitive configuration, called damping (α_G, α_R) and spectral shaping (S_G, S_R) spaces, respectively. The intuitive manipulation of α_G and α_R was achieved by a calibration process that consisted in determining a specific domain for each material category in the (α_G, α_R) space. Borders between material domains were determined based on results from predictive discrimination analysis. In practice, we calibrated the (α_G, α_R) space delimited by extreme values of α_G and α_R for typical sounds (range of $\alpha_G = [0.25; 3.34]$ and range of $\alpha_R = [0.5; 6.64] \times 10^{-4}$; see Fig. 7). This space was sampled in 2500 evenly spaced points

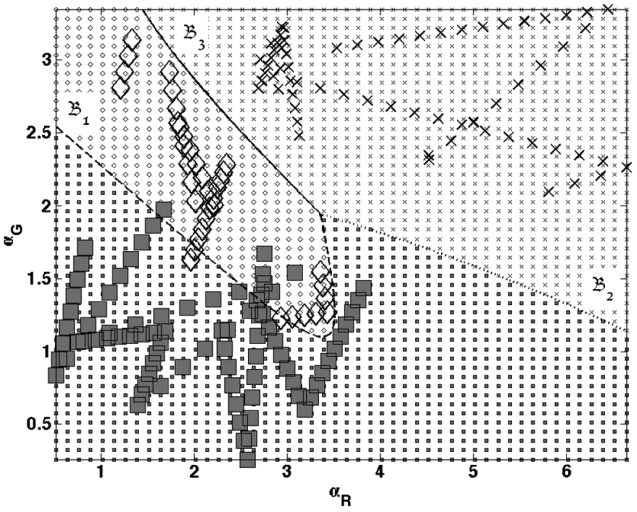


Fig. 7. Calibration of the $\{\alpha_G, \alpha_R\}$ space from border sections $\mathfrak{B}_1$ between Metal and Glass (dashed line), $\mathfrak{B}_2$ between Metal and Wood (dotted line) and $\mathfrak{B}_3$ between Glass and Wood (solid line). The positions of typical sounds for Wood ($\times$), Metal ($\square$), and Glass ($\diamond$) used for the classification process are also represented.

and each point was associated with a posterior probability of belonging to a material category. This probability was computed from a Bayesian rule based on the knowledge of the positions of typical sounds in the (α_G, α_R) space. Classification functions were determined between pairs of material categories and were expressed as quadratic combination of α_G and α_R ($\times 10^{-4}$). Boundary curves were materialized from the set of points that have similar classification probabilities δ for both categories

$$\{\mathbf{x} : \delta_{G_1}(\mathbf{x}) = \delta_{G_2}(\mathbf{x})\} \tag{23}$$

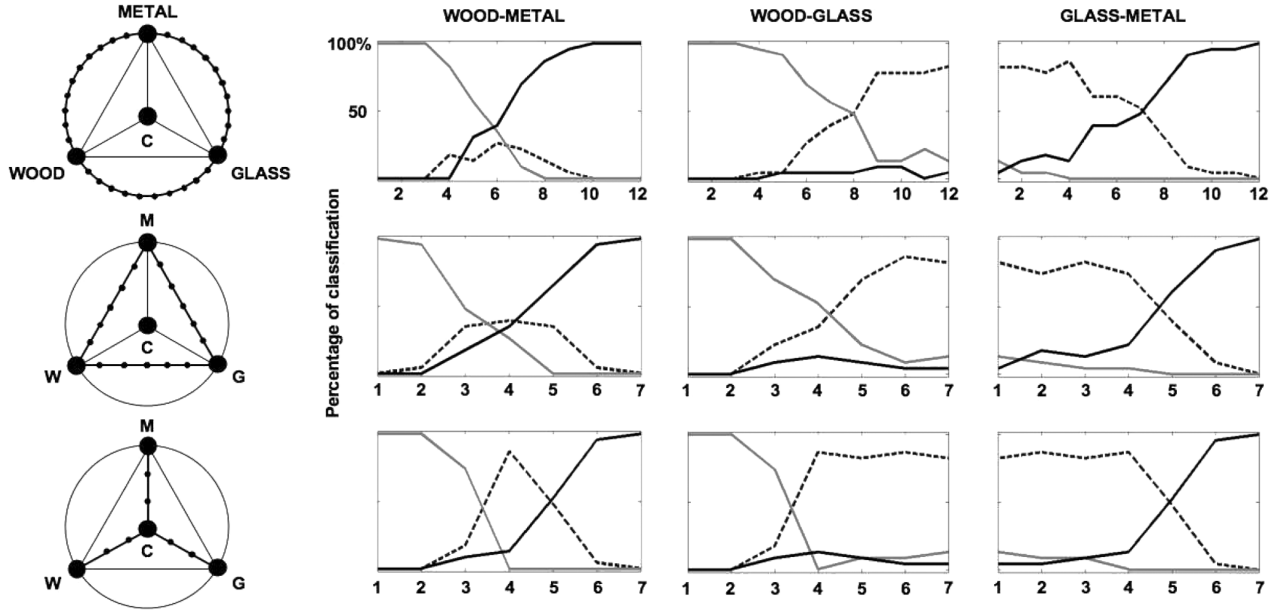


Fig. 8. Perceptual evaluation test of the control strategy. Left: position of sounds (black markers) used for the test in the Material Space as a function of the trajectory: along the external circle (first row), along the chords whose endpoints are the reference sounds (second row) and along the radii of the circle from the center C to the reference sounds (third row). Right: percentage of classification as Wood (gray curves), Metal (black curves), or Glass (dashed curves) corresponding to the three types of trajectory for each sound as a function of its position of the continuum.

or equivalently

$$\{\mathbf{x} : 0 = \delta_{G_1}(\mathbf{x}) - \delta_{G_2}(\mathbf{x})\} \tag{24}$$

where $\mathbf{x} = (\alpha_R, \alpha_G)$ and G_1 and G_2 the categories.

For our concern, the border noted $\mathfrak{B}_1$ between Metal and Glass categories was defined by

$$\mathfrak{B}_1 : 0 = 109.69 - 48.44\alpha_R - 48.30\alpha_G + 5.33(\alpha_R)^2 + 11.33\alpha_R\alpha_G + 3.37(\alpha_G)^2 \tag{25}$$

and all the points for which this function is negative were classified into the Metal category. The border $\mathfrak{B}_2$ between Wood and Metal was defined by

$$\mathfrak{B}_2 : 0 = -41.18 + 1.50\alpha_R + 17.75\alpha_G + 0.27(\alpha_R)^2 - 0.09\alpha_R\alpha_G - 0.20(\alpha_G)^2 \tag{26}$$

and all the points for which this function is negative were classified into the Wood category. Finally, the border $\mathfrak{B}_3$ between Wood and Glass regions was defined by

$$\mathfrak{B}_3 : 0 = 68.51 - 46.94\alpha_R - 30.55\alpha_G + 5.60(\alpha_R)^2 + 11.24\alpha_R\alpha_G + 3.17(\alpha_G)^2 \tag{27}$$

and all the points for which this function is negative were classified into the Wood category. The calibration of the (α_G, α_R) space was completed by keeping the section of the borders that directly separate the two sound categories: as shown in Fig. 7, the border section that was kept for $\mathfrak{B}_1$ is represented by a dashed line, the section kept for $\mathfrak{B}_2$ is represented by a dotted line and the section kept for $\mathfrak{B}_3$ is represented by a solid line. These borders were reported in the Middle layer (Fig. 6) allowing an intuitive manipulation of damping parameters. Note that these borders do not represent a strict delimitation between sound categories and a narrow transition zone may be taken into account on both sides of the borders. In particular, sounds belonging to this transition zone may be perceived as ambiguous sounds such as sounds created at intermediate positions of the continua.

Finally, the bottom layer concerns the signal parameters determined as follows: the damping coefficients α_m are computed from (4) with $\{\alpha_G, \alpha_R\}$ values and frequencies $\tilde{\omega}_m$ from (15) with $\{S_G, S_R\}$ values. The amplitudes A_m are assumed to be equal to one.

V. Perceptual Evaluation of the Control Strategy

The proposed control strategy for the perceived material was evaluated with a formal perceptual test. Twenty-three participants (9 women, 14 men) participated in the experiment. Sounds were selected in the Material Space as shown in Fig. 8 (left). Three types of trajectory between two reference sounds were investigated: along the external circle (by a 12-step continuum), along chords whose endpoints are the reference sounds (7-step continuum) and along the radii of the circle from the center C to the reference sounds (7-step continuum). Sounds were presented once randomly through headphones. The whole set of sounds are available at [38]. Participants were asked to categorize each sound as Wood, Metal, or Glass, as fast as possible, by selecting with a mouse on a computer screen the corresponding label. The order of labels displayed on the screen was balanced across participants. The next sound was presented after a 2-seconds silence. Participants' responses were collected and averaged for each category (Wood, Metal, and Glass) and for each sound.

Fig. 8 (right) shows results as a function of sound position along the continua. Sounds at extreme positions were classified by more than 70% of participants in the correct category, leading

to the validation of the reference sounds as typical exemplars of their respective material category. In Wood–Metal transition, sounds at intermediate positions were classified as Glass with highest percentages for those along the trajectory via the center C. By contrast, in both Wood–Glass and Glass–Metal transitions, intermediate sounds were most often classified in one of the two categories corresponding to the extreme sounds. From an acoustic point of view, this reflects the fact that, the interpolation of Damping parameters between Metal and Wood sounds crosses the Glass category while this is not the case for the other two transitions (see Fig. 6).

These results were in line with the ones obtained from the behavioral data in the first categorization experiment (Fig. 2) and consequently, allowed us to validate the proposed control strategy as an efficient way to navigate in the Material space. Note that the interpolation process was computed on a linear scale between parameters of the reference sounds [cf. (21) and (22)]. The next step will consist in taking into account these results and modify the interpolation rules so that the metric distance between a given sound and a reference sound in the Material space closely reflects perceptual distance.

VI. Conclusion

In this paper, we proposed a control strategy for the perceived material in an impact sound synthesizer. To this end, we investigated the sound characteristics relevant for an accurate evocation of material by conducting a sound categorization experiment. To design the stimuli, sounds produced by impacting three different materials, i.e., wood, metal, and glass, were recorded and synthesized by using analysis–synthesis techniques. After tuning, sound continua simulating progressive transitions between material categories were built and used in the categorization experiment. Both behavioral data and electrical brain activity were collected. From behavioral data, a set of typical sounds for each material category was determined and an acoustic analysis including descriptors known to be relevant for timbre perception and material identification was conducted. The most relevant descriptors that allow discrimination between material categories were identified: the normalized sound decay (related to the damping) and the roughness. Electrophysiological data provided complementary information regarding the perceptual/cognitive aspects related to the sound categorization and were discussed in the context of synthesis. Based on acoustic and ERP data, results confirmed the importance of damping and highlighted the relevance of spectral descriptors for material perception. Control strategies for damping and spectral shaping were proposed and tested in synthesis applications. These strategies were further integrated in a three-layer control architecture allowing the user to navigate in a "Material Space." A formal perceptual evaluation confirmed the validity of the proposed control strategy. Such a control offered an intuitive manipulation of parameters and allowed defining realistic impact sounds directly from the material label (i.e., Wood, Metal, or Glass).

References

[1] M. V. Mathews, "The digital computer as a musical instrument," *Science*, vol. 142, no. 3592, pp. 553–557, 1963.

[2] R. Moog, "Position and force sensors and their application to keyboards and related controllers," in *Proc. AES 5th Int. Conf.: Music Digital Technol.*, 1987, pp. 179–181, A. E. S. New York, Ed..

[3] M. Battier, "L'approche gestuelle dans l'histoire de la lutherie électronique. Etude d'un cas: Le theremin," *Proc. Colloque International* , ser. Collection Eupalinos, Editions Parenthèses, 1999, Les nouveaux gestes de la musique.

[4] J. Tenney, "Sound-generation by means of a digital computer," *J. Music Theory*, vol. 7, no. 1, Spring, 1963.

[5] A. Camurri, M. Ricchetti, M. Di Stefano, and A. Stroscio, "Eyesweb—Toward gesture and affect recognition in dance/music interactive systems," in *Proc. Colloquio di Informatica Musicale*, 1998.

[6] P. Gobin, R. Kronland-Martinet, G. A. Lagesse, T. Voinier, and S. Ystad, *From Sounds to Music: Different Approaches to Event Piloted Instruments*, ser. Lecture Notes in Computer Science. : Springer-Verlag, 2003, vol. 2771, pp. 225–246.

[7] M. Wanderley and M. Battier, "Trends in gestural control of music," IRCAM-Centre Pompidou, 2000.

[8] J.-C. Risset and D. L. Wessel, "Exploration of timbre by analysis and synthesis," in *The Psychology of Music*, ser. Cognition and Perception, 2nd ed. New York: Academic, 1999, pp. 113–169.

[9] D. L. Wessel, "Timbre space as a musical control structure," *Comput. Music J.*, vol. 3, no. 2, pp. 45–52, 1979.

[10] S. Ystad and T. Voinier, "A virtually-real flute," *Comput. Music J.*, vol. 25, no. 2, pp. 13–24, Summer, 2001.

[11] J. M. Grey, "Multidimensional perceptual scaling of musical timbres," *J. Acoust. Soc. Amer.*, vol. 61, no. 5, pp. 1270–1277, 1977.

[12] C. L. Krumhansl, "Why is musical timbre so hard to understand," in *Structure and Perception of Electroacoustic Sound and Music*. Amsterdam, The Netherlands: Elsevier, 1989.

[13] J. Krimphoff, S. McAdams, and S. Winsberg, "Caractérisation du timbre des sons complexes. II: Analyses acoustiques et quantification psychophysique [characterization of timbre of complex sounds. II: Acoustical analyses and psychophysical quantification]," *J. Phys.*, vol. 4, no. C5, pp. 625–628, 1994.

[14] S. McAdams, S. Winsberg, S. Donnadieu, G. D. Soete, and J. Krimphoff, "Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes," *Psychol. Res.*, vol. 58, pp. 177–192, 1995.

[15] W. A. Sethares, "Local consonance and the relationship between timbre and scale," *J. Acoust. Soc. Amer.* , vol. 94, no. 3, pp. 1218–1228, 1993.

[16] P. N. Vassilakis, "Auditory roughness as a means of musical expression," Dept. of Ethnomusicology, Univ. of California, Selected reports in ethnomusicology (perspectives in systematic musicology), 2005, vol. 12, pp. 119–144.

[17] R. P. Wildes and W. A. Richards, *Recovering Material Properties From Sound*, W. A. Richards, Ed. Cambridge, MA: MIT Press, 1988, ch. 25, pp. 356–363.

[18] W. W. Gaver, "How do we hear in the world ? explorations of ecological acoustics," *Ecol. Psychol.*, vol. 5, no. 4, pp. 285–313, 1993.

[19] R. Lutfi and E. Oh, "Auditory discrimination of material changes in a struck-clamped bar," *J. Acoust. Soc. Amer.*, vol. 102, no. 6, pp. 3647–3656, 1997.

[20] R. L. Klatzky, D. K. Pai, and E. P. Krotkov, "Perception of material from contact sounds," *Presence: Teleoperators and Virtual Environments*, vol. 9, no. 4, pp. 399–410, 2000.

[21] B. L. Giordano and S. McAdams, "Material identification of real impact sounds: Effects of size variation in steel, wood, and Plexiglas plates," *J. Acoust. Soc. Amer.*, vol. 119, no. 2, pp. 1171–1181, 2006.

[22] M. Aramaki, M. Besson, R. Kronland-Martinet, and S. Ystad, "Timbre perception of sounds from impacted materials: Behavioral, electrophysiological and acoustic approaches," in *Computer Music Modeling and Retrieval—Genesis of Meaning of Sound and Music*, ser. LNCS, S. Ystad, R. Kronland-Martinet, and K. Jensen, Eds. Berlin, Heidelberg, Germany: Springer-Verlag, 2009, vol. 5493, pp. 1–17.

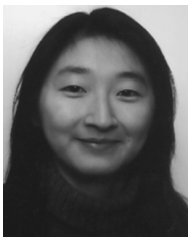
[23] M. Aramaki, R. Kronland-Martinet, T. Voinier, and S. Ystad, "A percussive sound synthesizer based on physical and perceptual attributes," *Comput. Music J.*, vol. 30, no. 2, pp. 32–41, 2006.

[24] M. Aramaki, R. Kronland-Martinet, T. Voinier, and S. Ystad, "Timbre control of a real-time percussive synthesizer," in *Proc. 19th Int. Congr. Acoust. (CD-ROM)*, 2007, 84-87985-12-2.

[25] M. Aramaki, C. Gondre, R. Kronland-Martinet, T. Voinier, and S. Ystad, "Thinking the sounds: An intuitive control of an impact sound synthesizer," in *Proc. 15th Int. Conf. Auditory Display (ICAD 2009)*, 2009.

[26] K. Steiglitz and L. E. McBride, "A technique for the identification of linear systems," *IEEE Trans. Autom. Control*, vol. AC-10, no. 10, pp. 461–464, Oct. 1965.

[27] R. Roy and T. Kailath, "Esprit-estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 7, pp. 984–995, Jul. 1989.
[28] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propagat.*, vol. AP-34, no. 3, pp. 276–280, Mar. 1986.
[29] R. Badeau, B. David, and G. Richard, "High-resolution spectral analysis of mixtures of complex exponentials modulated by polynomials," *IEEE Trans. Signal Process.*, vol. 54, no. 4, pp. 1341–1350, Apr. 2006.
[30] L.-M. Reissell and D. K. Pai, "High resolution analysis of impact sounds and forces," in *Proc. WHC '07: 2nd Joint EuroHaptics Conf. Symp. Haptic Interfaces for Virtual Environment and Teleoperator Syst.*, Washington, DC, 2007, pp. 255–260, IEEE Computer Society.
[31] M. Aramaki and R. Kronland-Martinet, "Analysis-synthesis of impact sounds by real-time dynamic filtering," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, no. 2, pp. 695–705, Mar. 2006.
[32] R. Kronland-Martinet, P. Guillemain, and S. Ystad, "Modelling of natural sounds by time-frequency and wavelet representations," *Organised Sound*, vol. 2, no. 3, pp. 179–191, 1997.
[33] R. Parncutt, *Harmony—A Psychoacoustical Approach*. Berlin/Heidelberg, Germany: Springer, 1989.
[34] A. Chaigne and C. Lambourg, "Time-domain simulation of damped impacted plates: I. Theory and experiments," *J. Acoust. Soc. Amer.*, vol. 109, no. 4, pp. 1422–1432, 2001.
[35] T. Ono and M. Norimoto, "Anisotropy of dynamic young's modulus and internal friction in wood," *Jpn. J. Appl. Phys.*, vol. 24, no. 8, pp. 960–964, 1985.
[36] S. McAdams, A. Chaigne, and V. Roussarie, "The psychomechanics of simulated sound sources: Material properties of impacted bars," *J. Acoust. Soc. Amer.*, vol. 115, no. 3, pp. 1306–1320, 2004.
[37] M. Aramaki, H. Baillères, L. Brancheriau, R. Kronland-Martinet, and S. Ystad, "Sound quality assessment of wood for xylophone bars," *J. Acoust. Soc. Amer.*, vol. 121, no. 4, pp. 2407–2420, 2007.
[38] 2010 [Online]. Available: http://www.lma.cnrs-mrs.fr/~kronland/Categorization/sounds.html, last checked: Oct. 2009
[39] M. Aramaki, L. Brancheriau, R. Kronland-Martinet, and S. Ystad, "Perception of impacted materials: Sound retrieval and synthesis control perspectives," in *Computer Music Modeling and Retrieval—Genesis of Meaning of Sound and Music*, ser. LNCS, S. Ystad, R. Kronland-Martinet, and K. Jensen, Eds. Berlin, Heidelberg, Germany: Springer-Verlag, 2009, vol. 5493, pp. 134–146.
[40] M. D. Rugg and M. G. H. Coles, "The ERP and cognitive psychology: Conceptual Issues," in *Electrophysiology of Mind. Event-Related Brain Potentials and Cognition*, ser. Oxford Psychology. New York: Oxford Univ. Press, 1995, pp. 27–39, no. 25.
[41] J. Eggermont and C. Ponton, "The neurophysiology of auditory perception: From single-units to evoked potentials," *Audiol. Neuro-Otol.*, vol. 7, pp. 71–99, 2002.
[42] A. Shahin, L. E. Roberts, C. Pantev, L. J. Trainor, and B. Ross, "Modulation of p2 auditory-evoked responses by the spectral complexity of musical sounds," *NeuroReport*, vol. 16, no. 16, pp. 1781–1785, 2005.
[43] S. Kuriki, S. Kanda, and Y. Hirata, "Effects of musical experience on different components of meg responses elicited by sequential pianotones and chords," *J. Neurosci.*, vol. 26, no. 15, pp. 4046–4053, 2006.
[44] E. Kushnerenko, R. Ceponiene, V. Fellman, M. Huotilainen, and I. Winkler, "Event-related potential correlates of sound duration: Similar pattern from birth to adulthood," *NeuroReport*, vol. 12, no. 17, pp. 3777–2781, 2001.
[45] C. Alain, B. M. Schuler, and K. L. McDonald, "Neural activity associated with distinguishing concurrent auditory objects," *J. Acoust. Soc. Amer.*, vol. 111, no. 2, pp. 990–995, 2002.
[46] S. McAdams, "Perspectives on the contribution of timbre to musical structure," *Comput. Music J.*, vol. 23, no. 3, pp. 85–102, 1999.
[47] H.-G. Kim, N. Moreau, and T. Sikora, *MPEG-7 Audio and Beyond: Audio Content Indexing and Retrieval*. New York: Wiley, 2005.
[48] G. Peeters, "A Large set of audio features for sound description (similarity and description) in the Cuidado Project," IRCAM, Paris, France, 2004, Tech. Rep..
[49] J. Marozeau, A. de Cheveigné, S. McAdams, and S. Winsberg, "The dependency of timbre on fundamental frequency," *J. Acoust. Soc. Amer.*, vol. 114, pp. 2946–2957, 2003.
[50] J. W. Beauchamp, "Synthesis by spectral amplitude and "brightness" matching of analyzed musical instrument tones," *J. Audio Eng. Soc.*, vol. 30, no. 6, pp. 396–406, 1982.
[51] R. Plomp and W. J. M. Levelt, "Tonal consonance and critical bandwidth," *J. Acoust. Soc. Amer.*, vol. 38, pp. 548–560, 1965.
[52] E. Terhardt, "On the perception of periodic sound fluctuations (roughness)," *Acustica*, vol. 30, no. 4, pp. 201–213, 1974.
[53] P. N. Vassilakis, "SRA: A web-based research tool for spectral and roughness analysis of sound signals," in *Proc. 4th Sound Music Comput. (SMC) Conf.*, 2007, pp. 319–325.
[54] 2009 [Online]. Available: http://www.cycling74.com/, last checked: Octob. 2009
[55] D. Rocchesso and F. Fontana, 2003, "The Sounding Object," [Online]. Available: http://www.soundobject.org/SobBook/SobBook_JUL03.pdf last checked: Oct. 2009
[56] M. Hyde, "The N1 response and its applications," *Audiol. Neuro-Otol.*, vol. 2, pp. 281–307, 1997.
[57] C. Valette and C. Cuesta, *Mécanique de la corde vibrante (Mechanics of vibrating string)*, ser. Traité des Nouvelles Technologies, série Mécanique. London, U.K.: Hermès, 1993.
[58] N. H. Fletcher and T. D. Rossing, *The Physics of Musical Instruments (Second Edition)*. Berlin, Germany: Springer-Verlag, 1998.
[59] E. Zwicker and H. Fastl, *Psychoacoustics, Facts and Models*. Berlin, Germany: Springer-Verlag, 1990.
[60] H. L. F. von Helmholtz, *On the Sensations of Tone as the Physiological Basis for the Theory of Music*, 2nd ed. New York: Dover, 1877, reprinted 1954.



Mitsuko Aramaki (M'09) received the M.S. degree in mechanics (speciality in acoustic and dynamics of vibrations) from the University of Aix-Marseille II, Marseille, France, and the Ph.D. degree for her work at the Laboratoire de Mécanique et d'Acoustique, Marseille, France, in 2003, on analysis and synthesis of impact sounds using physical and perceptual approaches.

She is currently a Researcher at the Institut de Neurosciences Cognitives de la Méditerranée, Marseille, where she works on a pluridisciplinary project combining sound modeling, perceptual and cognitive aspects of timbre, and neuroscience methods in the context of virtual reality.

Mireille Besson received the Ph.D. degree in neurosciences from the University of Aix-Marseille II, Marseille, France, in 1984.

After four years of post-doctorate studies at the Department of Cognitive Science, University of California at San Diego, La Jolla, working with Prof. M. Kutas and at the Department of Psychology, University of Florida, Gainesville, working with Prof. I. Fischler, she obtained a permanent position at the National Center for Scientific Research (CNRS), Marseille, France. She is currently Director of Research at the CNRS, Institut de Neurosciences Cognitives de la Méditerranée (INCM), where she is the head of the "Language, Music, and Motor" team. Her primary research interests are centered on brain imaging of linguistic and non linguistic sound perception and on brain plasticity mainly using event-related brain potentials. She is currently conducting a large research project on the influence of musical training on linguistic sound perception in normal reading and dyslexic children.

Richard Kronland-Martinet (M'09–SM'10) received the M.S. degree in theoretical physics in 1980, the Ph.D. degree in acoustics from the University of Aix-Marseille II, Marseille, France, in 1983, and the "Doctorat d'Etat es Sciences" degree from the University of Aix-Marseille II in 1989 for his work on analysis and synthesis of sounds using time–frequency and time–scale (wavelets) representations.

He is currently Director of Research at the National Center for Scientific Research (CNRS), Laboratoire de Mécanique et d'Acoustique, Marseille, where he is the Head of the group "Modeling, Synthesis and Control of Sound and Musical Signals." His primary research interests are in analysis and synthesis of sounds with a particular emphasis on high-level control of synthesis processes. He recently addressed applications linked to musical interpretation and semantic description of sounds using a pluridisciplinary approach associating signal processing, physics, perception, and cognition.

Sølvi Ystad received the Ph.D. degree in acoustics from the University of Aix-Marseille II, Marseille, France, in 1998.

She is currently a Researcher at the National French Research Center (CNRS) in the research team S2M—Synthesis and Control of Sounds and Musical Signals—in Marseille, France. Her research activities are related to sound modeling with a special emphasis on the identification of perceptually relevant sound structures to develop efficient synthesis models. She was in charge of the research project "Towards the sense of sounds," financed by the French National Agency (ANR—http://www.sensons.cnrs-mrs.fr) from 2006–2009.

PLOS ONE

RESEARCH ARTICLE

Seeing Circles and Drawing Ellipses: When Sound Biases Reproduction of Visual Motion

Etienne Thoret[1]*, Mitsuko Aramaki[1], Lionel Bringoux[2], Sølvi Ystad[1], Richard Kronland-Martinet[1]

1 Laboratoire de Mécanique et d'Acoustique, CNRS, UPR 7051, Aix Marseille Université, Centrale Marseille, Marseille, France, **2** Aix-Marseille Université, CNRS, ISM, UMR 7287, Marseille, France

* etienne.thoret@mcgill.ca

CrossMark
click for updates

OPEN ACCESS

Citation: Thoret E, Aramaki M, Bringoux L, Ystad S, Kronland-Martinet R (2016) Seeing Circles and Drawing Ellipses: When Sound Biases Reproduction of Visual Motion. PLoS ONE 11(4): e0154475. doi:10.1371/journal.pone.0154475

Editor: Susana Martinez-Conde, State University of New York Downstate Medical Center, UNITED STATES

Received: December 2, 2015

Accepted: April 14, 2016

Published: April 27, 2016

Copyright: © 2016 Thoret et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This work was funded by the French National Research Agency (ANR) under the MetaSon: Métaphores Sonores (Sound Metaphors) project (ANR-10-CORD-0003) in the CONTINT 2010 framework and the SoniMove Project (ANR-14-CE24-0018). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

The perception and production of biological movements is characterized by the 1/3 power law, a relation linking the curvature and the velocity of an intended action. In particular, motions are perceived and reproduced distorted when their kinematics deviate from this biological law. Whereas most studies dealing with this perceptual-motor relation focused on visual or kinaesthetic modalities in a unimodal context, in this paper we show that auditory dynamics strikingly biases visuomotor processes. Biologically consistent or inconsistent circular visual motions were used in combination with circular or elliptical auditory motions. Auditory motions were synthesized friction sounds mimicking those produced by the friction of the pen on a paper when someone is drawing. Sounds were presented diotically and the auditory motion velocity was evoked through the friction sound timbre variations without any spatial cues. Remarkably, when subjects were asked to reproduce circular visual motion while listening to sounds that evoked elliptical kinematics without seeing their hand, they drew elliptical shapes. Moreover, distortion induced by inconsistent elliptical kinematics in both visual and auditory modalities added up linearly. These results bring to light the substantial role of auditory dynamics in the visuo-motor coupling in a multisensory context.

Introduction

It is now well established that biological motion is characterized by specific kinematic properties, for instance, by the 1/3 power law which postulates that the tangential velocity of the motion v_t is constrained by the local curvature C of its geometrical trajectory: $v_t = KC^{-1/3}$ with K a constant [1]. Regarding the visual modality, it has been shown that the human ability to track a visual motion with the non-hidden hand is facilitated when the motion complies with the 1/3 power law [2]. By contrast, the perceived geometry of a circular visual motion can be distorted if the motion does not comply with these biological rules [3]. More recently, we showed that the visuo-motor coupling of circular motions displayed with incongruent elliptical kinematics were distorted by subjects who do not see their hand [4]. Regarding the kinaesthetic modality, Viviani, Baud-Bovy, and Redolfi [5] have also shown that the perception of the movement geometry is constrained by the covariations between the movement kinematics and

Competing Interests: The authors have declared that no competing interests exist.

its curvature. Indeed, a circular hand movement trailed by a mechanical arm that does not comply with the 1/3-power law distorts its perceived geometry into an elliptical one. Taken together, these results confirmed the existing relation between curvature and dynamics in the emergence of the perceived and/or reproduced motion geometry.

In the field of multisensory research, many studies have demonstrated the ability of the auditory modality to modify the visual perception of motion [6,7] (see [8] for a review) and even drive our motor behavior [9] (see [10] for a review). For instance, Brooks et al. [11] showed that the visual perception of a specific kind of biological movement, represented by a point-like walker, is affected by spatial auditory motion. In line with this observation, Arrighi et al. [12] demonstrated that sounds can even enhance the visual perception of point-like walkers when displayed with synchronous auditory motions. These studies give credit to a multisensory processing framework of biological motions.

The present study investigated the influence of movement dynamics on the reproduced geometry of biological motion by combining auditory and visual stimuli evoking circular or elliptical motions in consistent or inconsistent audiovisual situations. Whereas spatial (i.e. geometrics-related) cues are intrinsically conveyed by visual (and kinaesthetic) sensory inputs, we hereby investigated purely dynamic (i.e., kinematics-related) cues without spatial information by using monophonic sounds [13]. Indeed, recent experiments revealed that time-varying, monophonic friction sounds produced by someone drawing evoke the movement kinematics of the drawer's pencil and enable, to a certain extent, the recognition of the drawn shape [14]. In our experiment, subjects were asked to synchronize drawing movements on a graphic tablet with visual motions displayed on a screen without seeing their hand. It was hypothesized that sounds evoking incongruent kinematics with respect to the visual motion affect the geometry of the reproduced shape resulting from the coupling between a visual biological motion and a drawing movement. This would highlight the role of movement dynamics in the emergence of the geometry independently from spatial cues.

By investigating the effect of congruent and incongruent combinations of auditory and visual kinematics on visuo-motor coupling of biological motion, we aimed at considering crossmodal influences of these two modalities. As previously assumed, if movement dynamics alone influence the geometry of the motion induced by such a visuo-motor coupling, it would allow for the investigation of the combined role of the auditory and visual channels in a multisensory context.

Materials & Methods

Participants

Seventeen right-handed subjects (2 women; M = 28.5 years, SD = 8 years) participated in the experiment. All subjects provided their consent prior to the study and were naive as to the specific purpose of the experiment. At the time the experiment was designed and conducted (winter 2012–2013), no ethics approval was required from the Aix-Marseille University for behavioral studies such as those reported in this manuscript. The local Ethics board at the Aix-Marseille University subsequently approved highly similar experiments conducted in the same Institute two years later the experiment reported in this manuscript was conducted. Neither of the experiments involved deception or stressful procedures. Participants were informed that they were free to leave the experiment at any time, and that their data would have been treated anonymously. The research reported in this manuscript was carried out according to the principles expressed in the 1964 Declaration of Helsinki. Participants in the experiment were recruited on a voluntary basis from the students and staff of the Groupement des Laboratoires de Marseille. All the data were anonymized before analysis.

Stimuli

Audiovisual stimuli evoking motions were generated from specific kinematic rules applied to both visual and auditory stimuli leading to consistent or inconsistent situations. According to the kinematic behavior of planar motion, the dynamic behavior was modeled by a system of two harmonic oscillators of period T that differ by their relative phase, noted Φ, allowing for the simulation of specific kinematic properties ranging from circular to elliptical motion [13]:

$$\begin{cases} x(s(t)) = Amp \cdot \cos\left(\frac{2\pi}{T}s(t)\right) \\ y(s(t)) = Amp \cdot \cos\left(\frac{2\pi}{T}s(t) + \Phi\right) \end{cases}$$

where *Amp* is the amplitude of the motion, x and y the coordinates of the motion in the (x(t), y (t)) plane, s(t) the curvilinear abscissa along the trajectory, and t the time. Interestingly, this model complies with biological rules, in particular the 1/3 power law. Two configurations were used here: 1) circular kinematics, with relative phase Φ = 90°, leading to a constant velocity profile matching the kinematic behavior associated with the displacement along a circle and 2) elliptical kinematics, with Φ = 45°, leading to a time varying velocity profile matching the kinematics associated with the displacement along an ellipse of eccentricity 0.9, which corresponds to the ellipse drawn in the most natural way [15].

Visual motions consisted of a white moving dot, with a diameter of 6mm on a black background. The motion was generated according to a method proposed by Viviani, Baud-Bovy, and Redolfi [5]. The dot always followed the same trajectory, a geometric circle with a radius (R) of 6.36 cm corresponding to a circumference of 40 cm, while its velocity varied according to either the circular or the elliptic kinematics as described above (Fig 1, Left). They defined what is referred to as the Visual Kinematics (VK) and were characterized by the relative phase, noted Φ_V (indexed V for visual modality). For circular kinematics obtained with $\Phi_V = 90°$ and A = R corresponding to constant velocity, the kinematics were consistent with biological

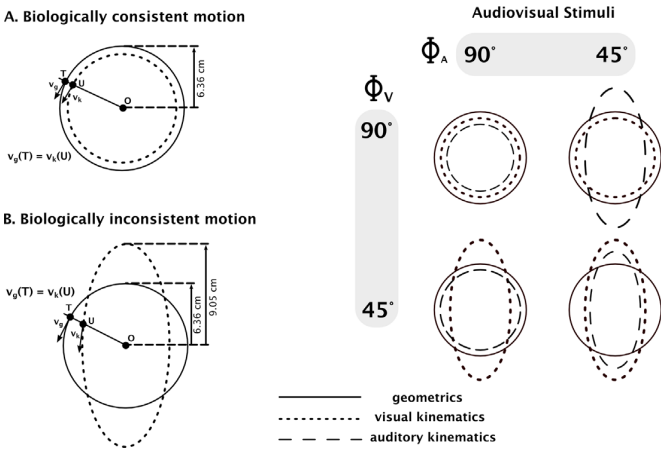


Fig 1. Visual Stimuli. Left: Kinematics of visual stimuli (dotted line) that either comply (Panel A) or not (Panel B) with biological rules for a geometric circle (solid line). The velocity of the dot in the biologically consistent motion condition is constant and equals 22.24 cm.s^{-1} and varies between 13.81 cm.s^{-1} and 31.62 cm.s^{-1} for the biologically inconsistent condition. **Right:** Audiovisual stimuli consisting of combinations of visual and auditory motion characterized by their relative phases Φ (indexed V and A for vision and audition respectively).

doi:10.1371/journal.pone.0154475.g001

motion on the circular geometric trajectory traveled by the dot (Fig 1, Left, Panel A). For elliptical kinematics, with $\Phi_V = 45°$ and A = 6.92 cm corresponding to a time varying velocity, the dot moved faster in the vertical parts of the circle and the motion was therefore biologically inconsistent (Fig 1 Left, Panel B), with the circular trajectory traveled by the dot. Each type of visual motion contained 19 complete cycles of period T = 1.8s and lasted for 34.2s. The four audiovisual stimuli are available in S1–S4 Videos.

Auditory motion was evoked by synthesized friction sounds generated according to a physically-informed model [16,17], which considers that a friction sound is the result of successive micro-impacts produced when a sharp object (exciter) interacts with the asperities of a rough surface (resonator). From a signal point of view, such friction sounds can be simulated by bandpass filtered white noise, in which the central filter frequency is controlled by the velocity profile. A biquad bandpass filter with a constant quality factor equal to 3 was used for this purpose. The relationship between the tangential velocity v_t and the central frequency f_c of the bandpass filter was defined by: $f_c(t) = \alpha\, v_t(t)$ where α is a constant proportionality coefficient. From a perceptual point of view, the mapping mainly influences the timbre, in particular the mean spectral centroid of the resulting friction sound: the higher the value of α, the higher the mean spectral centroid (perceived brightness). The most appropriate mapping (chosen as α = 20) was based on results from a preliminary calibration experiment. For the sake of realism of the synthetic-sound, the contribution of an interacting resonant object was taken into account, and the sounds were synthesised to evoke rubbing on a wooden plate [18,19]. Based on this model, synthesized friction sounds modulated by the velocity profile corresponding either to circular or elliptical kinematics were generated. They defined what is referred to as the Auditory Kinematics (AK) and were characterized by the relative phase, noted Φ_A (indexed A for auditory modality). For circular kinematics ($\Phi_A = 90°$), the generated sound contained no modulation and evoked a uniform motion, whereas for elliptical kinematics ($\Phi_A = 45°$), the generated sound contained acoustic modulations induced by the time-varying velocity profile. These modulations reflect variations of specific acoustical features, such as the spectral centroid that has been shown to adequately evoke perceived motion [20–22].

Hence, 12 different types of audiovisual motion (AVM) were obtained by generating four combinations of constant and time-varying, visual and auditory motions (4 conditions x 3 repetitions, see Fig 1, Right). The AVM had consistent kinematics when both visual and auditory kinematics matched uniform circular kinematics ($\Phi_V = \Phi_A = 90°$). In this case, the AVM corresponded to biological motion. In the three other cases, the AVM had inconsistent kinematics with respect to the geometry as elliptical kinematics were conveyed either in the auditory modality ($\Phi_V = 90°$; $\Phi_A = 45°$), in the visual modality ($\Phi_V = 45°$; $\Phi_A = 90°$) or in both modalities ($\Phi_V = \Phi_A = 45°$). These three types of AVM were therefore non biological.

Task

Subjects were placed in front of a DELL 1907FP screen (1280 x 1024 pixels; 60 Hz refresh rate) that displayed the moving dot while they listened to the auditory stimuli, presented diotically without any spatial cues through a set of headphones (Sennheiser HD650 headphones, the sample rate of the soundcard was 44100 Hz with 16-bit resolution), and were asked to synchronize their hand movement with the visual moving dot on a graphic tablet (Wacom Intuos 5 graphic tablet at a sample rate of 129 Hz and with a spatial precision of 5.10^{-3} mm; see Fig 2). No specific information was given concerning the auditory kinematics or the geometry of the trajectory to be reproduced. In particular, the subjects did not know that the visual trajectory of the moving dot was always circular and that only the kinematics of the moving dot differed. The stimuli were presented according to two different pseudo random series that were

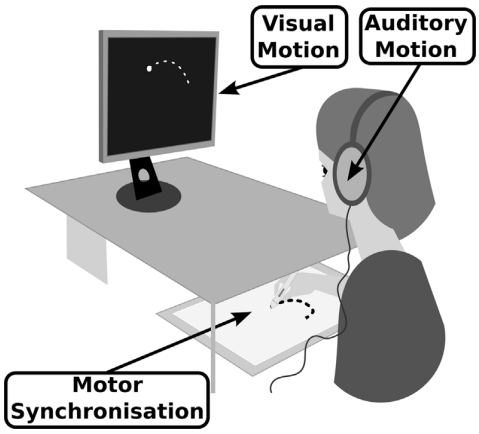


Fig 2. Experimental set-up. The subjects had to synchronize their hand movement with the moving dot on a graphic tablet without seeing their drawing hand.

doi:10.1371/journal.pone.0154475.g002

balanced across subjects: CC CE EE EE CE EC CC CE EE CC EC CE and EC CC CE EC EC CC EE CE CE EE EE CC, the audiovisual conditions are denoted as follow CC ($\Phi_V = 90°$; $\Phi_A = 90°$), CE ($\Phi_V = 90°$; $\Phi_A = 45°$), EC ($\Phi_V = 45°$; $\Phi_A = 90°$), and EE ($\Phi_V = 45°$; $\Phi_A = 45°$). A training session was conducted before the experiment to familiarise subjects with the task and with the use of the graphic tablet. During the training session, which contained up to 3 trials of the actual test, the same instructions as in the real test were given. In order to evaluate the visuo-motor coupling without visual feedback on the performed movement, the subjects could not see their hand during the experiment. In order to analyze the influence of the different types of AVM on motor performance, the drawn shapes were fitted to compute their relative phases Φ_{drawn}.

Data Analyses

Data collected on the graphic tablet were filtered using a Savitzky-Golay filter [23] with a 43-point temporal window and third-order interpolation to remove digital noise due to the high sampling rate. Then high-pass Butterworth filtering (0.2 Hz cut-off frequency) was applied to remove the spatial drift observed on hand movements (due to the fact that subjects could not see their hand while drawing). Finally, data were analysed with respect to the relative phase Φ_{drawn} characterising the geometry of the reproduced shape. First, the eccentricity e_{drawn} (i.e. a variable characterising the flatness of the drawn shape) was estimated for the last 10 out of 19 drawn shapes [4,5] and then the relative phase Φ_{drawn} was calculated using the following formula: $\Phi_{drawn} = 2\arctan\sqrt{1 - e_{drawn}{}^2}$.

The statistical design contained 2 Visual Kinematics (VK) x 2 Auditory Kinematics (AK). Repeated measures ANOVA was performed with Statistica software to evaluate the effects of each experimental factor on the motor performance described by the relative phase Φ_{drawn}. The normality and the homogeneity of the distributions were assessed with a Lilliefors test and a O'Brien test respectively. The distortion between the reproduced shapes and the circular geometry of the visual motion in terms of flatness was performed by means of a one-sample two-tailed t-test between the relative phase Φ_{drawn} and the mean $\Phi_V = 90°$ in the four audiovisual conditions. The significance level of the p-value was set to 0.05 for all analyses.

Table 1. Results of the experiment.

Φ_V	90°		45°	
Φ_A	90°	45°	90°	45°
Φ_{drawn}	77.6	72.1	65.9	60.6
	[76.1 79.0]	[68.3 75.9]	[63.6 68.3]	[57.03 64.31]
e_{drawn}	0.594	0.685	0.761	0.811
	[0.566 0.622]	[0.625 0.734]	[0.734 0.784]	[0.777 0.839]

Means and 95% Confidence Intervals of the Relative Phases and Eccentricities Characterizing the Drawn Shapes for the Four Conditions.

doi:10.1371/journal.pone.0154475.t001

Results

The averaged relative phases by subjects are within S1 Dataset.

The results showed that the geometry of the drawn shapes was noticeably distorted by elliptical kinematics conveyed by both visual and auditory modalities (see Table 1). More precisely a main effect of the visual kinematics ($F(1,16) = 147.22$, $p < .001$) as well as a main effect of the auditory kinematics ($F(1,16) = 12.83$, $p < .01$) on the flatness, characterized by the relative phase of the reproduced shape was revealed, meaning that circles are reproduced flattened (i.e. elliptical) when the visual kinematics ($\Phi_V = 45°$) as well as auditory kinematics ($\Phi_A = 45°$) are elliptical. However, the interaction between visual and auditory kinematics was not significant ($F(1,16) = 0.01$, $p = 0.96$; Fig 3 and Fig 4). The S1–S4 Videos display the 4 averaged reproduced ellipses along with the corresponding audiovisual stimuli.

Figs 3 and 4 show that even in the consistent situation when both auditory and visual kinematics are circular, the reproduced circles were distorted into a more elliptical shape ($\Phi_{drawn} = 77°$). Globally, elliptical visual kinematics ($\Phi_V = 45°$) induced stronger distortions than elliptical auditory kinematics ($\Phi_A = 45°$). A distortion of 11.7° with respect to the congruent situation was observed when elliptical visual kinematics is combined with circular auditory kinematics ($\Phi_V = 45°$, $\Phi_A = 90°$), while a distortion of 5.5° is observed when elliptical auditory kinematics

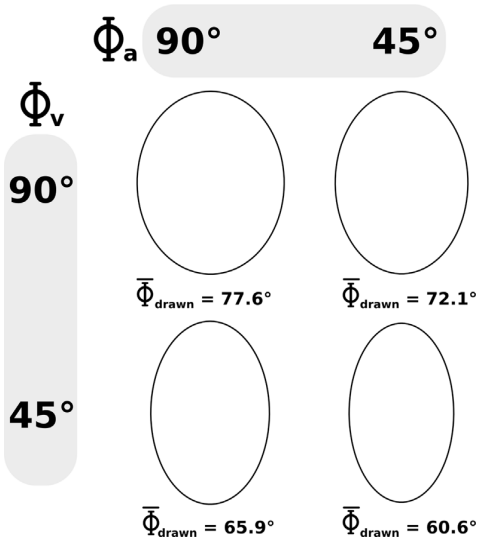


Fig 3. Motor performances. Mean of the motor reproduction in the four audiovisual conditions.

doi:10.1371/journal.pone.0154475.g003

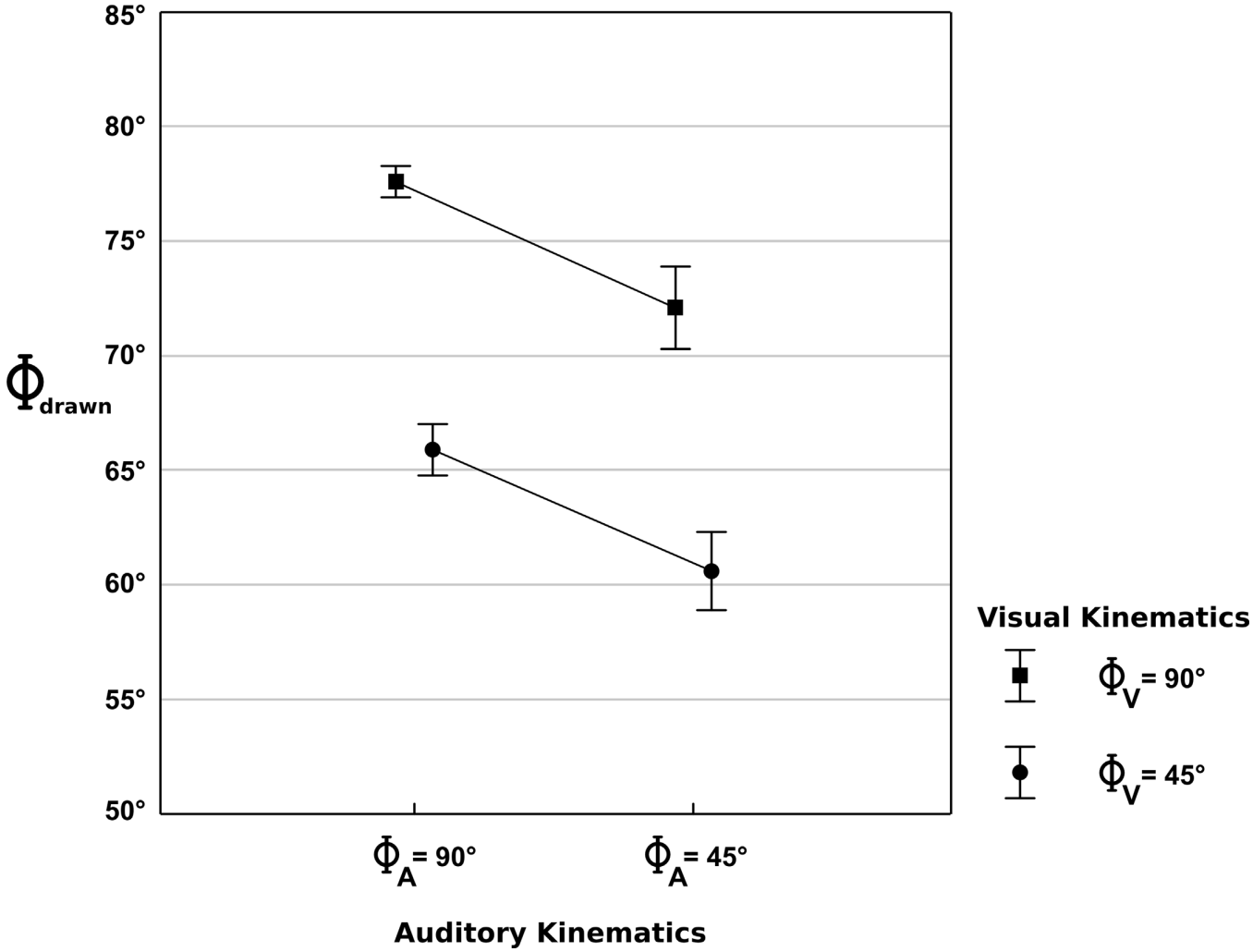


Fig 4. Results. Mean relative phase and 95% confidence intervals of the reproduced motion as a function of visual and auditory kinematics, illustrating the distinct influences of both factors (i.e., no interaction).

doi:10.1371/journal.pone.0154475.g004

is combined with circular visual kinematics ($\Phi_V = 90°$, $\Phi_A = 45°$). When both visual and auditory kinematics were elliptical ($\Phi_V = 45°$, $\Phi_A = 45°$), a distortion of 17° was obtained, which corresponds to the sum of the previously observed distortions obtained when either the visual or the auditory kinematics were elliptical. Hence, the distortion induced by elliptical auditory kinematics has not been influenced by visual kinematics (5.3° for $\Phi_V = 45°$ and 5.5° for $\Phi_V = 90°$) and vice versa (11.5° for $\Phi_A = 45°$ and 11.7° for $\Phi_A = 90°$). Finally, the reproduced motion was distorted in all situations (Two-tailed t-tests between Φ_{drawn} and 90°: $\Phi_V = 90°$, $\Phi_A = 90°$: $t(16) = -18.09$, $p < .001$; $\Phi_V = 90°$, $\Phi_A = 45°$: $t(16) = -9.89$, $p < .001$; $\Phi_V = 45°$, $\Phi_A = 45°$: $t(16) = -21.37$, $p < .001$; $\Phi_V = 45°$, $\Phi_A = 90°$: $t(16) = -17.06$, $p < .001$).

Discussion

In this study we investigated the influence of both visual and auditory dynamics on the reproduced geometry of visual motion by asking subjects to synchronize drawing movements on a

graphic tablet. The results were analyzed in terms of relative phase distortion of the reproduced shape with respect to the visual (circular) trajectory displayed on a screen in front of the subjects.

First, it was found that for all of the audiovisual conditions, the motor reproductions were significantly flatter than the actual circular motions displayed on the screen. This observation is in line with studies on movement coordination showing that when someone draws a repetitive enclosed shape like a circle, he/she naturally tends to draw an elliptical shape which may be considered as an attractor of such a dynamic system [15,24,25].

The results showed that elliptical visual kinematics flattened the reproduced circles regardless of the auditory kinematics. This effect of biologically inconsistent visual kinematics on the reproduced circle underlined the links between motor output and visual motion perception, confirming the results and expectations of Viviani and colleagues [2,3,5]. Our results, issued from a situation without visual feedback from the drawing movement, thereby showed that the motor reproduction of visual biological motion is constrained by co-variations between curvature and kinematic properties in a feedforward manner, which could not have been fully demonstrated in the visual feedback condition investigated by Viviani et al. [2].

However, it must be noted that we cannot conclude whether both visual perception and motor processes are affected by incongruent auditory motion as our experiments were performed in a multisensory and motor context. Nevertheless, we recently showed [5] that distortions induced by elliptical visual kinematics ($\Phi_V = 45°$) are larger for the visuo-motor coupling (10.17%) than the perceptual distortions observed by Viviani and colleagues in purely visual (1.01% in [3]) or kinaesthetic presentations (<1% in [5]). As we here observed that the motor reproduction was less distorted by auditory elliptical kinematics ($\Phi_A = 45°$) than elliptical visual kinematics ($\Phi_V = 45°$), we may expect that if distorted, the perception of a visual circular motion ($\Phi_V = 90°$) would be less affected by elliptical auditory kinematics ($\Phi_A = 45°$) than by elliptical visual kinematics ($\Phi_V = 45°$).

The key point of this study is that continuous friction sounds also substantially interfered with visual motion cues and clearly modified the geometric properties of the motor reproduction by flattening the reproduced circles regardless of the visual kinematics in the biologically inconsistent condition. In line with our hypothesis, this effect of sound on the reproduced geometry strongly attests to the central role of dynamics *per se*, on the visuo-motor coupling. While the influence of visual kinematics involves both geometric and kinematic cues, the influence of auditory dynamics evoked by timbre variations of friction sounds allows for the investigation of the role of kinematic cues alone and reveal that the emergence of the geometry and the associated motor output is clearly driven by motion dynamics. Note that this would not have been possible to investigate the role of auditory dynamics with rhythmic discrete sounds, e.g. corresponding to velocity minima or maxima, as such stimuli convey only the timing of the movement and not its continuous velocity variations. Hence, while the studies made by Viviani et al. [2,3,5] supporting perceptual-motor interactions were performed in a purely unimodal context, in the present study we showed that such interactions also can occur in a multisensory context. In line with this, Varlet et al. [26] investigated sensorimotor coordination between audio-visual motion and a motor synchronization. They reported that continuous sounds affect motor synchronization. Although their study did not focus on the geometry but on temporal aspects of motor synchronization with audiovisual stimuli on a horizontal axis, their results highlighted the influence of motor attractors and biomechanical constraints in the sensori-motor coupling with audiovisual motions.

The absence of interaction between visual and auditory kinematics in our study suggests that the visual and auditory influences were combined in a perfectly linear way. While both the visual and auditory motions were inconsistent, i.e. elliptical ($\Phi_V = 45°$, $\Phi_A = 45°$), the distortion of the reproduced shape equaled the sum of the individual distortions obtained when either the

visual motion was biologically inconsistent ($\Phi_V = 45°$, $\Phi_A = 90°$) or the auditory motion was inconsistent ($\Phi_V = 90°$, $\Phi_A = 45°$). Although the literature on multisensory integration most often claims that the effects of several sensory channels add up non-linearly and that synchronous sensory inputs amplify or inhibit perceptual and behavioral effects (e.g., [27]), some behavioral studies did not support this view and revealed additive effects [28–31]. Here, we provide a clear example of linear integration between visual and auditory motion-related cues. More than the additivity of the effects observed here, it is noticeable that the distortion induced by biologically inconsistent kinematics in the visual domain is stronger (about 12°) than the distortion induced by inconsistent auditory kinematics (about 5°). This suggests that the effect of dynamics per se is about half the combined effect of spatial and dynamic cues.

Taken together, these results could be interpreted in line with the Theory of Event Coding [32] stressing the existence of unified percepts that combine both sounds produced by objects and modulated by actions with their perceptual properties. Thoret et al. [13] suggested the existence of such a unified percept of biological movements from a purely auditory point of view, i.e. linking the properties of drawing movements to their evocation through timbre variations of friction sounds. Studies by Danna et al. [33] confirmed such binding between friction sounds and drawing movements by showing that the quality of handwriting can be accurately judged through friction sounds evoking the movement velocity. Such unified percepts that link sounds and actions have also been shown for footstep sounds [34]. Our results are clearly in line with the existence of this kind of unified percept in a multisensory context involving audio-visuo-motor coupling. In particular, the motor production appears to be driven by the linear integration of both geometric and dynamic cues in the visual modality and solely by dynamic cues evoked through timbre variations in the auditory modality.

Finally, these findings may help in designing new devices using sounds to enhance or substitute visual and, more generally, sensory information. In particular, training activities involving continuous auditory stimuli when treating diseases that affect motor function, such as dysgraphia [35] and Parkinson's disease [36,37], or when guiding movements in sport [38], are believed to constitute powerful alternatives to existing methods. In this context, we suggest that conveying dynamic information through specific sound properties can substantially modify motor performance.

Conclusions

The experiment presented in this study provides an evidence of the central role of dynamics in the reproduction of the geometry of a visual motion in an audiovisual context. We showed that both biologically inconsistent visual motion and inconsistent auditory motion flatten the reproduced geometry of circular visual motions. Interestingly, the combined effects of visual and auditory inputs observed here were added up linearly. Nevertheless, in order to better understand these effects and how the two modalities are combined to plan motor actions, it might be interesting to manipulate the instructions given to the subjects. This could be done by asking subjects either to focus on the sound rather than the visual motion, or to focus with the same attention on the visual and auditory information. The experimental data could be analyzed in terms of sensorimotor coordination such as in the study of Varlet et al. [26]. The relative phase considered in our study as a geometrical descriptor characterizing the eccentricity of the ellipse could then be used to reveal the temporal coordination of the movements in terms of negative versus non-negative lag and stability, which might also be affected by incoherent audiovisual motions. Moreover, purely perceptual experiments investigating whether incongruent audiovisual motions affect the visual perception of the geometry would enable to reveal information on the separate role of perceptual and motor processes in such a task.

In a more general perspective, it might be of interest to evaluate whether the effects observed here are specific to biological motions or whether the motor reproduction of any kind of audio-visual motion might be affected by incongruent visual and auditory kinematics, for instance by considering physical motions such as those constrained by Newton's laws [39].

Supporting Information

S1 Video. Audiovisual stimulus ($\Phi_V = 90°$; $\Phi_A = 90°$) and the corresponding averaged reproduced movement. The visual sample rate of the original visual stimulus has been down-sampled to 30 Hz and the audio encoded in the AAC format due to technical capabilities.
(MP4)

S2 Video. Audiovisual stimulus ($\Phi_V = 90°$; $\Phi_A = 45°$) and the corresponding averaged reproduced movement. The visual sample rate of the original visual stimulus has been down-sampled to 30 Hz and the audio encoded in the AAC format due to technical capabilities.
(MP4)

S3 Video. Audiovisual stimulus ($\Phi_V = 45°$; $\Phi_A = 90°$) and the corresponding averaged reproduced movement. The visual sample rate of the original visual stimulus has been down-sampled to 30 Hz and the audio encoded in the AAC format due to technical capabilities.
(MP4)

S4 Video. Audiovisual stimulus ($\Phi_V = 45°$; $\Phi_A = 45°$) and the corresponding averaged reproduced movement. The visual sample rate of the original visual stimulus has been down-sampled to 30 Hz and the audio encoded in the AAC format due to technical capabilities.
(MP4)

S1 Dataset. Averaged relative phases by subjects.
(TXT)

Acknowledgments

The authors are thankful to Meghan Goodchild for revising the English of the manuscript.

Author Contributions

Conceived and designed the experiments: ET MA LB SY RKM. Performed the experiments: ET. Analyzed the data: ET MA LB SY RKM. Contributed reagents/materials/analysis tools: ET MA LB SY RKM. Wrote the paper: ET MA LB SY RKM.

References

1. Lacquaniti F, Terzuolo C, Viviani P (1983) The law relating the kinematic and figural aspects of drawing movements. Acta Psychologica 54: 115–130. doi: 10.1016/0001-6918(83)90027-6 PMID: 6666647
2. Viviani P, Campadelli P, Mounoud P (1987) Visuo-manual pursuit tracking of human two-dimensional movements. Journal of Experimental Psychology: Human Perception and Performance 13: 62–78. doi: 10.1037//0096-1523.13.1.62
3. Viviani P, Stucchi N (1989) The effect of movement velocity on form perception: Geometric illusions in dynamic displays. Perception & Psychophysics 46: 266–274. doi: 10.3758/bf03208089
4. Thoret E, Aramaki M, Bringoux L, Ystad S, Kronland-Martinet R (2016) When eyes drive hand: Influence of non-biological motion on visuo-motor coupling. Neuroscience Letters 612: 225–230. doi: 10.1016/j.neulet.2015.12.022 PMID: 26708633
5. Viviani P, Baud-Bovy G, Redolfi M (1997) Perceiving and tracking kinesthetic stimuli: Further evidence of motor-perceptual interactions. Journal of Experimental Psychology: Human Perception and Performance 23: 1232–1252. doi: 10.1037/0096-1523.23.4.1232 PMID: 9269735

6. Kitagawa N, Ichihara S (2002) Hearing visual motion in depth. Nature 416: 172–174. doi: 10.1038/416172a PMID: 11894093
7. Sekuler R, Sekuler A, Lau R (1997) Sound alters visual motion perception. Nature 385: 308–308. doi: 10.1038/385308a0 PMID: 9002513
8. Soto-Faraco S, Kingstone A, Spence C (2003) Multisensory contributions to the perception of motion. Neuropsychologia 41: 1847–1862. doi: 10.1016/s0028-3932(03)00185-4 PMID: 14527547
9. Repp B, Penel A (2003) Rhythmic movement is attracted more strongly to auditory than to visual rhythms. Psychological Research 68. doi: 10.1007/s00426-003-0143-8
10. Repp B, Su Y (2013) Sensorimotor synchronization: A review of recent research (2006–2012). Psychonomic Bulletin & Review 20: 403–452. doi: 10.3758/s13423-012-0371-2
11. Brooks A, van der Zwan R, Billard A, Petreska B, Clarke S, Blanke O (2007) Auditory motion affects visual biological motion processing. Neuropsychologia 45: 523–530. doi: 10.1016/j.neuropsychologia.2005.12.012 PMID: 16504220
12. Arrighi R, Marini F, Burr D (2009) Meaningful auditory information enhances perception of visual biological motion. Journal of Vision 9: 25–25. doi: 10.1167/9.4.25
13. Thoret E, Aramaki M, Kronland-Martinet R, Velay JL, Ystad S (2014) From sound to shape: Auditory perception of drawing movements. Journal of Experimental Psychology: Human Perception and Performance 40: 983–994. doi: 10.1037/a0035441 PMID: 24446717
14. Hollerbach J (1981) An oscillation theory of handwriting. Biol Cybern 39: 139–156. doi: 10.1007/bf00336740
15. Dounskaia N, Van Gemmert A, Stelmach G (2000) Interjoint coordination during handwriting-like movements. Experimental Brain Research 135: 127–140. doi: 10.1007/s002210000495 PMID: 11104134
16. Gaver WW (1993) How Do We Hear in the World? Explorations in Ecological Acoustics. Ecological Psychology 5: 285–313. doi: 10.1207/s15326969eco0504_2
17. van den Doel K, Kry P, Pai D (2001) FoleyAutomatic: Physically-Based Sound Effects for Interactive Simulation and Animation. Proceedings of the 28th annual conference on Computer graphics and interactive techniques—SIGGRAPH '01. doi: 10.1145/383259.383322
18. Conan S, Thoret E, Aramaki M, Derrien O, Gondre C, Ystad S, et al. (2014) An Intuitive Synthesizer of Continuous-Interaction Sounds: Rubbing, Scratching, and Rolling. Computer Music Journal 38: 24–37. doi: 10.1162/comj_a_00266
19. Aramaki M, Besson M, Kronland-Martinet R, Ystad S (2011) Controlling the Perceived Material in an Impact Sound Synthesizer. IEEE Transactions on Audio, Speech, and Language Processing 19: 301–314. doi: 10.1109/tasl.2010.2047755
20. Chowning J (1977) The Simulation of Moving Sound Sources. Computer Music Journal 1: 48–52. doi: 10.2307/3679609
21. Kronland-Martinet R, Voinier T (2008) Real-Time Perceptual Simulation of Moving Sources: Application to the Leslie Cabinet and 3D Sound Immersion. EURASIP Journal on Audio, Speech, and Music Processing 2008: 849696. doi: 10.1186/1687-4722-2008-849696
22. Merer A, Aramaki M, Ystad S, Kronland-Martinet R (2013) Perceptual characterization of motion evoked by sounds for synthesis control purposes. ACM Transactions on Applied Perception 10: 1–24. doi: 10.1145/2422105.2422106
23. Savitzky A, Golay M (1964) Smoothing and Differentiation of Data by Simplified Least Squares Procedures. Analytical Chemistry 36: 1627–1639. doi: 10.1021/ac60214a047
24. Athènes S, Sallagoïty I, Zanone P, Albaret J (2004) Evaluating the coordination dynamics of handwriting. Human Movement Science 23: 621–641. doi: 10.1016/j.humov.2004.10.004 PMID: 15589625
25. Danna J, Athènes S, Zanone P (2011) Coordination dynamics of elliptic shape drawing: Effects of orientation and eccentricity. Human Movement Science 30: 698–710. doi: 10.1016/j.humov.2010.08.019 PMID: 21524807
26. Varlet M, Marin L, Issartel J, Schmidt R, Bardy B (2012) Continuity of Visual and Auditory Rhythms Influences Sensorimotor Coordination. PLoS One 7: e44082. doi: 10.1371/journal.pone.0044082 PMID: 23028488
27. McGarry L, Russo F, Schalles M, Pineda J (2012) Audio-visual facilitation of the mu rhythm. Exp Brain Res 218: 527–538. doi: 10.1007/s00221-012-3046-3 PMID: 22427133
28. Angelaki D, Gu Y, DeAngelis G (2009) Multisensory integration: psychophysics, neurophysiology, and computation. Current Opinion in Neurobiology 19: 452–458. doi: 10.1016/j.conb.2009.06.008 PMID: 19616425

29. Campos J, Byrne P, Sun H (2010) The brain weights body-based cues higher than vision when estimating walked distances. European Journal of Neuroscience 31: 1889–1898. doi: 10.1111/j.1460-9568.2010.07212.x PMID: 20584194

30. Campos J, Butler J, Bülthoff H (2012) Multisensory integration in the estimation of walked distances. Exp Brain Res 218: 551–565. doi: 10.1007/s00221-012-3048-1 PMID: 22411581

31. Campos J, Butler J, Bülthoff H (2014) Contributions of visual and proprioceptive information to travelled distance estimation during changing sensory congruencies. Exp Brain Res 232: 3277–3289. doi: 10.1007/s00221-014-4011-0 PMID: 24961739

32. Hommel B, Müsseler J, Aschersleben G, Prinz W (2001) The Theory of Event Coding (TEC): A framework for perception and action planning. Behavioral and Brain Sciences 24: 849–878. doi: 10.1017/s0140525x01000103 PMID: 12239891

33. Danna J, Paz-Villagrán V, Gondre C, Aramaki M, Kronland-Martinet R, Ystad S, et al. (2015) "Let Me Hear Your Handwriting!," Evaluating the Movement Fluency from Its Sonification. PLoS One 10: e0128388. doi: 10.1371/journal.pone.0128388 PMID: 26083384

34. Young W, Rodger M, Craig C (2013) Perceiving and reenacting spatiotemporal characteristics of walking sounds. Journal of Experimental Psychology: Human Perception and Performance 39: 464–476. doi: 10.1037/a0029402 PMID: 22866760

35. Danna J, Fontaine M, Paz-Villagrán V, Gondre C, Thoret E, Aramaki M, et al. (2015) The effect of real-time auditory feedback on learning new characters. Human Movement Science 43: 216–228. doi: 10.1016/j.humov.2014.12.002 PMID: 25533208

36. Rodger M, Young W, Craig C (2014) Synthesis of Walking Sounds for Alleviating Gait Disturbances in Parkinson's Disease. IEEE Trans Neural Syst Rehabil Eng 22: 543–548. doi: 10.1109/tnsre.2013.2285410 PMID: 24235275

37. Young W, Rodger M, Craig C (2014) Auditory observation of stepping actions can cue both spatial and temporal components of gait in Parkinson's disease patients. Neuropsychologia 57: 140–153. doi: 10.1016/j.neuropsychologia.2014.03.009 PMID: 24680722

38. Craig C, Delay D, Grealy M, Lee D (2000). Guiding the swing in golf putting. Nature 405: 295–296. doi: 10.1038/35012690 PMID: 10830947

39. La Scaleia B, Zago M, Moscatelli A, Lacquaniti F, Viviani P (2014). Implied dynamics biases the visual perception of velocity. PloS One, 9(3), e93020. doi: 10.1371/journal.pone.0093020 PMID: 24667578

Simon Conan, Etienne Thoret, Mitsuko Aramaki, Olivier Derrien, Charles Gondre, Sølvi Ystad, and Richard Kronland-Martinet
Laboratoire de Mécanique et d'Acoustique
CNRS, UPR 7051
31 chemin Joseph Aiguier
13402 Marseille Cedex 20, France
{conan, thoret, aramaki, derrien, gondre, ystad, kronland}@lma.cnrs-mrs.fr

An Intuitive Synthesizer of Continuous-Interaction Sounds: Rubbing, Scratching, and Rolling

Abstract: In this article, we propose a control strategy for synthesized continuous-interaction sounds. The framework of our research is based on the *action–object* paradigm that describes the sound as the result of an action on an object and that presumes the existence of sound invariants (i.e., perceptually relevant signal morphologies that carry information about the action's or the object's attributes). Auditory cues are investigated here for the evocations of rubbing, scratching, and rolling interactions. A generic sound-synthesis model that simulates these interactions is detailed. We then suggest an intuitive control strategy that enables users to navigate continuously from one interaction to another in an "action space," thereby offering the possibility to simulate morphed interactions—for instance, ones that morph between rubbing and rolling.

Synthesis of everyday sounds is still a challenge, especially the control of sound-synthesis processes. Indeed, it is of interest to intuitively control sounds obtained with a synthesis model, that is, to be able to create sounds that carry or evoke specific information. To achieve this, we need to offer users the possibility to create and control sounds from semantic descriptions of sound events or from gestures. Intuitive sound-synthesis control provides interesting alternatives to indexed sound databases in domains such as the development of video games (Lloyd, Raghuvanshi, and Govindaraju 2011; Böttcher 2013), and is of great interest for sound design (Farnell 2010), sonification (Dubus and Bresin 2013), and virtual and augmented reality—for instance, for motor rehabilitation (Danna et al. 2013; Rodger, Young, and Craig 2013).

In previous studies, intuitive control of sounds based on acoustic descriptors or features has been proposed through so-called feature synthesis (Hoffman and Cook 2006). Other authors either have suggested that sounds can be directly generated from semantic descriptions of timbre (Gounaropoulos and Johnson 2006; Le Groux and Verschure 2008), or have evoked motion of the sound source (Merer et al. 2013). Aramaki et al. (2006, 2009a) proposed the use of semantic labels describing the perceived material, size, and shape of the object producing a sound to intuitively control an impact-sound synthesizer. In particular, that control strategy allowed the user to apply a continuous control to the perceived material to simulate continuous transitions (i.e., morphing) from one material to another (e.g., from glass to metal, through a continuum of ambiguous materials).

This article is devoted to the synthesis and control of continuous-interaction sounds. By continuous interaction we mean any kind of friction phenomena (Akay 2002) or rolling. We look at a subset of continuous interactions—i.e., rubbing, scratching, and rolling sounds. Similarly to the continuous control space of perceived material offered by Aramaki and co-workers (2009a), we would like to present a control space for actions that enables continuous transitions, for instance, from rubbing to rolling. From a synthesis point of view, we have a generic model that allows for such continuous sound transformations. We have achieved this by investigating previous synthesis models: some based on physical modeling or physically informed considerations (Gaver 1993; Hermes 1998; van den Doel, Kry, and Pai 2001; Rath and Rocchesso 2004; Stoelinga and Chaigne 2007) and others on analysis-synthesis schemes (Lagrange, Scavone, and Depalle 2010; Lee, Depalle, and Scavone 2010). Such a generic tool is of interest for sound design and for fundamental studies in sound perception (Aramaki et al. 2009b; Micoulaud-Franchi et al. 2011).

Computer Music Journal, 38:4, pp. 24–37, Winter 2014
doi:10.1162/COMJ_a_00266
© 2014 Massachusetts Institute of Technology.

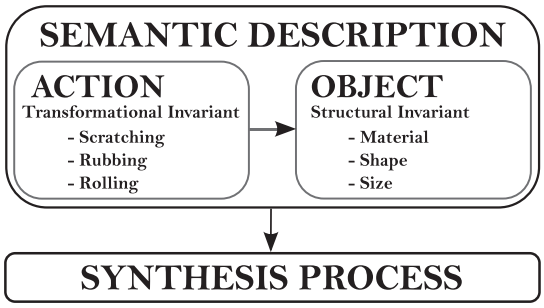


Figure 1. The action–object framework for the synthesis and control of sounds.

In the next section, we describe the *action–object* paradigm within which we developed our synthesizer. We then examine sound morphologies that convey the evocation of rubbing, scratching, and rolling interactions, and describe how these interactions can be reproduced by synthesis. Further, a control strategy of the proposed synthesis model is presented. In the last section, we provide some general conclusions and outline future research.

The Action–Object Paradigm

Inspired by William Gaver (1993), who proposed an independent synthesis of actions and objects, we developed a conceptual description of sounds through an action–object paradigm. This concept consists of considering the sounds as resulting from an action on a resonant object—e.g., "plucking a metal string" or "hitting a wooden plate." This approach suggests the existence of specific acoustic patterns in the perceived signal, enabling the auditory identification of objects, on the one hand, and actions, on the other. For instance, a vibrating string produces a particular spectral content that enables the listener to recognize it, whether it is bowed (e.g., violin), plucked (e.g., guitar), or struck (e.g., piano). Similarly, it is possible to recognize a bottle by the sound it produces, whether it bounces or breaks (Warren and Verbrugge 1984), or a bouncing or rolling cylinder, and the material it is made of (Lemaitre and Heller 2012).

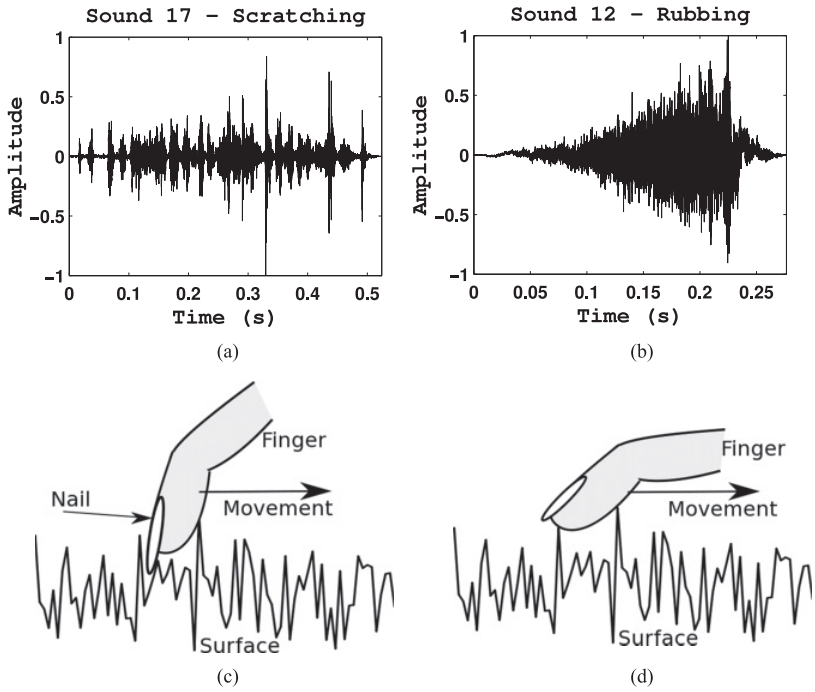
The psychological theory underlying this paradigm is known as the ecological approach of perception, first introduced by James Gibson for visual perception (Gibson 1966, 1979; for a more accessible introduction to Gibson's theory, refer to Michaels and Carello 1981). He proposed that the perception of visual events is constrained by our interactions with the surrounding world, and, more precisely, that the recognition of the properties of a visual event is provided by invariant structures contained in the sensory flow. Concerning the auditory recognition of acoustic events, this theory was first exploited by Warren and Verbrugge (1984) and later formalized by McAdams and Bigand (1993). It supposes the existence of *invariant structures* that carry the necessary information for the recognition of sound events. The so-called invariants are split in two categories: *structural invariants*, which enable recognition of physical properties of a sounding object (its material, shape, etc.) and *transformational invariants* describing the type of change or the action on the object (breaking, rolling, etc.).

Some studies have already identified such acoustic invariants. For instance, it has been shown that impact sounds contain sufficient information to perceptually discriminate the material (Wildes and Richards 1988) or the size (Lakatos, McAdams, and Caussé 1997; Carello, Anderson, and Kunkler-Peck 1998) of the sound-producing, impacted objects. In particular, it was shown that the perception of material is mainly related to frequency-dependent damping of spectral components (Klatzky, Pai, and Krotkov 2000; Tucker and Brown 2002; Giordano and McAdams 2006) and to roughness (Aramaki et al. 2009b). A study by Warren and Verbrugge (1984) revealed that, from the rhythm of a series of impact sounds, it is possible to predict if a glass will break or bounce. More recently, Thoret and colleagues (2014) highlighted that, by listening to friction sounds produced when someone is drawing, subjects were able to recognize (to a certain extent) the shape that was drawn and that the relevant information was conveyed by the velocity profile of the writer's gesture. The general action–object framework, on which the synthesis model presented in this article is based, is illustrated in Figure 1.

Invariants Related to Continuous Solid Interactions

In this section, acoustic invariants related to rubbing, scratching, and rolling interactions

Figure 2. Recorded sounds 100 percent associated with scratching (a) and with rubbing (b), as well as schematic representations of a nail scratching a surface (c) and a finger rubbing a surface (d). In the schematic representations, the y*-axis represents the detailed surface height (greatly exaggerated for clarity); the* x*-axis represents the direction along which the finger is dragged across the surface (also corresponding approximately to time, as the finger is dragged from left to right).*



are described. Technical details related to the implementation of both the object and the action parts will be provided later.

Rubbing and Scratching Interactions

To our knowledge, the auditory ability to distinguish rubbing sounds from scratching sounds has not previously been formally investigated. To ascertain that it is possible to distinguish a sound that evokes rubbing from one that evokes scratching and to reveal the signal properties responsible for this ability, a series of perceptual experiments was conducted by Conan and colleagues (2012; the related paper and experimental material are available online at www.lma.cnrs-mrs.fr/~kronland/RubbingScratching). We summarize the main results here. From the first experiment, qualitative analyses were made of recorded sounds for which all subjects agreed that the sound evoked one of the interactions, scratching or rubbing. These analyses implied that rubbing sounds resulted from a higher temporal density of impact events than scratching sounds. A sound that was associated with scratching by all subjects is plotted in Figure 2a, and a sound that was associated with rubbing by all subjects is in Figure 2b. The differences in temporal density of impacts can be explained as follows. Sounds produced when scratching a surface, for instance, with a nail, are due to the interaction between surface irregularities and the nail and therefore can be considered to be the result of successive impacts. Scratching a surface can be considered as scanning deeper into the surface than rubbing, which implies that each surface irregularity is encountered one after another and more intensely than in the case of rubbing. In rubbing, several surface irregularities are encountered simultaneously and less intensely, implying a more noisy sound, due to the higher density of impacts (see Figures 2c and 2d). The perceived surface irregularities therefore seem to be closely linked to the type of interaction.

These observations led us to set up a second experiment to validate the previous hypothesis. The experiment investigated how impact density, as a

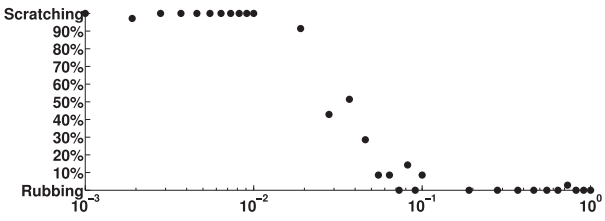


Figure 3. Results of the experiment with synthesized sounds. The x*-axis represents the temporal density of impacts (log scale, increasing from left to right), and the* y*-axis represents the percentage of association to the scratching category for each sound.*

relevant acoustic cue, could serve to distinguish rubbing from scratching sounds. The stimuli used for the experiment were synthetic friction sounds generated with different impact densities. The synthesis model was based on the pioneering work of Gaver (1993), relying on phenomenological considerations, and later improved by van den Doel, Kry, and Pai (2001). It consists of simulating the interaction force as a result of successive "microimpacts" of a plectrum on the irregularities of a surface. The successive impacts are modeled by low-pass filtered noise with a cutoff frequency related to the plectrum velocity, and the roughness of the surface is defined by the nature of the noise. Thirty synthetic sounds, representing a continuous transition from low- to high-density impact series, were generated. These sounds were then presented randomly to 35 subjects who were asked to associate each sound with one of the two interaction categories, rubbing or scratching. The results of the perceptual experiment are displayed in Figure 3 (details provided in Conan et al. 2012) and confirm our hypothesis that low impact densities are associated with scratching and high impact densities with rubbing, and there is a less clear perception of the type of interaction for moderate densities (at intermediary positions of the continuous transition).

In summary, these experiments allowed us to conclude that one invariant (possibly among others) that contributes to the discrimination between these two interactions is the temporal density of the impacts contained in the interaction force. That is, the higher the impact density in the signal, the higher the probability for the sound to evoke rubbing. Similarly, lower impact density in the signal tends to evoke scratching. Therefore, from a synthesis point of view, rubbing and scratching interactions can be simulated by controlling the characteristics of the impact series.

Rolling Interaction

The rolling interaction has been fully investigated by Conan and co-workers (2014). The main points are summarized in this section.

Similarly to rubbing and scratching interactions, we address here the determination of acoustic morphologies related to the auditory perception of rolling: Which signal information is responsible for the recognition of a rolling object? To answer this question, we investigated a physics-based model of rolling sounds. In the literature, most authors consider that the physics of a rolling ball is similar to the physics of a bouncing ball. The model of a bouncing ball generally takes into account a nonlinear sphere–plane interaction that relates the force f applied to the sphere to the penetration x and the penetration velocity $\dot{x}$ of the sphere into the surface (following the model of Hunt and Crossley 1975):

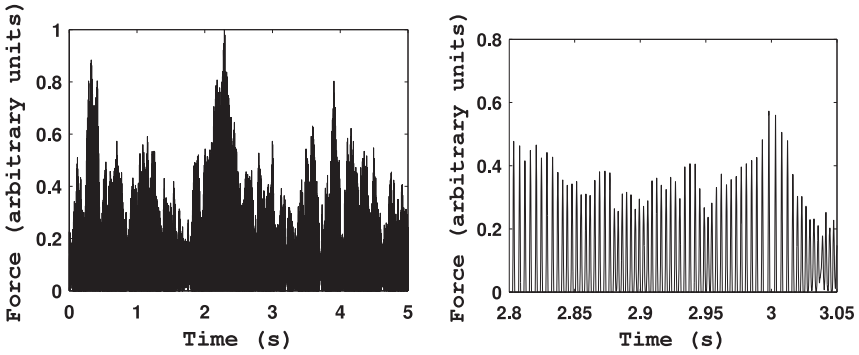
$$f(x,\dot{x}) = \begin{cases} kx^{\alpha} + \lambda x^{\alpha}\dot{x}, & x > 0 \\ 0, & x \leq 0, \end{cases} \tag{1}$$

where k is the stiffness and λ the damping weight of the force. The parameter α takes into account the local geometry around the contact surface ($\alpha = 3/2$ according to Hertz's theory of contact mechanics, cf. Johnson 1987). By taking into account the effect of the gravity on the ball, this model simulates the behavior of a bouncing ball (see, for instance, Falcon et al. 1998; Avanzini and Rocchesso 2001).

To adapt this model to the simulation of a rolling ball (Rath and Rocchesso 2005; Stoelinga and Chaigne 2007) or a rolling wheel (Nordborg 2002), consider that the rolling object moves along an irregular surface and the height of the irregularities is added as a perturbation to the penetration term x in Equation 1. The rolling interaction can be considered as a ball that bounces on surface irregularities with a randomly changing height.

We simulated the rolling of a ball over an irregular surface using the fourth-order Runge-Kutta method

Figure 4. Simulated rolling interaction force (a) and a magnified (zoomed in) excerpt (b). The force parameters are $\alpha = 3/2$, $k = 10^7 N/m^{3/2}$, $\lambda = 10^7 N\text{-}sec/m^{5/2}$. The ball has a mass of 5g and a velocity of 0.5m/sec. The surface is assumed to be fractal with $\beta = 1.2$ and a maximum amplitude of $10^{-9} m$.



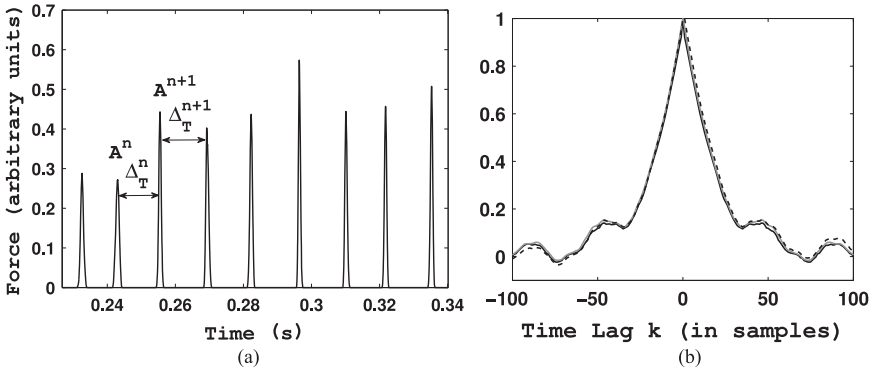
(a) (b)

(as described by Papetti, Avanzini, and Rocchesso 2011, where numerical issues of Equation 1 are studied). Note that we do not consider here the vibration of the surface itself, as do Rath and Rocchesso (2005). The surface, therefore, is modeled by noise with a specific spectrum, adjusted according to tribological observations (Ben Abdelounis et al. 2010) and phenomenological considerations (van den Doel, Kry, and Pai 2001; Rocchesso and Fontana 2003). In practice, the spectrum is characterized by $S(\omega) \propto 1/\omega^{\beta}$ where β enables the control of the perceived roughness. The amplitude of $S(\omega)$ is normalized to provide a maximum asperity of $10^{-9} m$. Also called fractal noise, such a spectrum accurately models most physical surfaces. An example of a simulated nonlinear interaction force f with such a surface is plotted in Figure 4.

Based on results from informal listening tests, such an interaction force f was found to convey sufficient information to evoke a rolling object (and was often perceived as a small, hard marble ball). Moreover, it has been shown that this model produces rolling sounds spontaneously recognized as such by naive listeners (Rath 2004). From a signal point of view, this force can be considered as a series of impacts (see Figure 4b). This assumption has already been exploited by Dik Hermes (1998) for the purpose of sound synthesis, and by Lagrange and colleagues (2010) in an analysis-synthesis context. In particular, if we consider the amplitude of the impacts and the time interval between two successive impacts as time series, the autocorrelation and the cross-correlation between the previously simulated time series obtained from the interaction force f can be computed. As shown in Figure 5b, these time series have a strong autocorrelation. That is, successive impacts have strong mutual dependencies. They are also strongly cross-correlated. These observations are coherent with the physics of a bouncing ball (recall that the rolling model is derived from a bouncing model), since within a bounce event, successive impacts are mutually related. Hence, a first conclusion that can be made is that the temporal structure of impact series associated with a rolling interaction seems to follow a specific pattern.

Another important characteristic of the rolling interaction force f is the dependency of the contact time on the impact velocity, which is related to the amplitude of the impact. This dependency has been studied by several authors (Chaigne and Doutaut 1997; Avanzini and Rocchesso 2001) and seems to be an important auditory cue that is responsible for evoking the rolling interaction. Indeed, to informally test this assumption, we first detected all the impact durations and amplitudes in the simulated force plotted in Figure 4. We then created two modified versions of this force by replacing the impact windows by a raised cosine window (van den Doel, Kry, and Pai 2001) that fits the original impact shapes well, one with a duration depending on the impact amplitude, the other with a fixed duration. It was always found that the version with varying, amplitude-dependent impact durations clearly produced the most realistic evocations

Figure 5. Notations for the impacts amplitudes A *and time interval* Δ_T *series (a). Autocorrelation of* Δ_T *(black) and* A *(gray) time series, cross-correlation between* A *and* Δ_T *time series (dashed black) (b). The amplitudes of the correlations functions are normalized.*



(a) (b)

of the rolling interaction, confirming previous findings.

Finally, some authors have suggested that a rolling object should never be considered perfectly spherical or perfectly homogeneous, so they have applied a sinusoidal modulation to the interaction force (Hermes 1998; Rath and Rocchesso 2005). From a perceptual point of view, the asymmetry of rolling objects is likely to contribute to the sensation of velocity. This auditory cue has been shown to strongly influence the perception of rolling objects' size and speed (Houben 2002). Hence, we considered amplitude modulation as a significant sound effect for evoking the sound of rolling.

Summary of Invariants Related to Continuous Interactions

The previous considerations allowed us to conclude that the interaction forces associated with rubbing, scratching, and rolling interactions can all be represented as impact series. Based on results from listening tests, it was shown that the temporal density of impacts conveys the information needed to recognize rubbing or scratching sounds. As opposed to rubbing and scratching interactions, the intrinsic structure of the impact series (correlations and statistics of durations and amplitudes) seems to be an indispensable signal morphology for evoking the rolling interaction. In the next section, we will describe a generic model to simulate rubbing, scratching, and rolling sounds based on the reproduction of the characteristics of these impact series.

Implementation of the Generic Sound-Synthesis Model

As described in the previous section, the interaction force carries relevant perceptual information related to rubbing, scratching, and rolling interactions. For these three types of interaction, the forces are series of impact events with specific relations. Such a signal can be formally described as:

$$f(t) = \sum_n A^n \, \phi^n(t - T^n), \tag{2}$$

where A^n and T^n are, respectively, the amplitude and time-position of each impact, and ϕ^n represents the "impact pattern," i.e., the shape of the n^{th} impact. Because the absolute time position of the impact is not a relevant parameter, we consider the time interval between successive impacts, defined as $\Delta_T^n = T^{n+1} - T^n$, in the remainder of this article (cf. Figure 5).

The aim of this section is to describe a synthesis process that is generic enough to simulate rubbing, scratching, and rolling interactions and, further, to design an intuitive control allowing a continuous navigation between these interactions. According to the action–object paradigm, the interaction forces constitute the action part of the general action–object framework, in which actions and

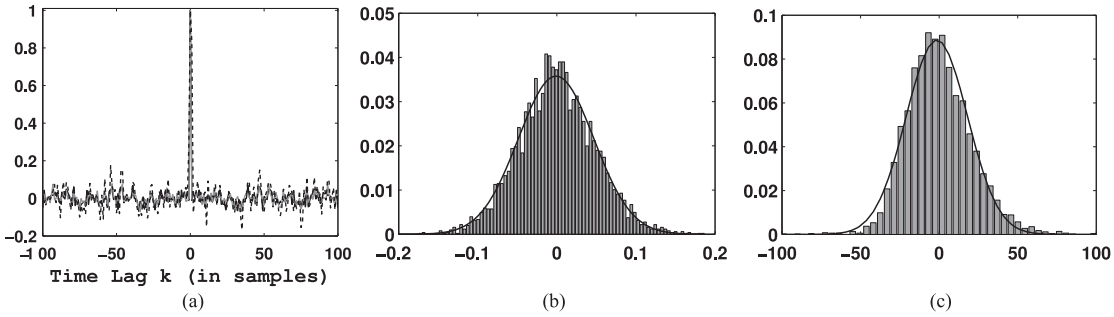


Figure 6. Autocorrelation of ${}^{c}\tilde{\Delta}_{T}$ (black) and ${}^{c}\tilde{A}$ (gray) time series, cross-correlation between ${}^{c}\tilde{A}$ and ${}^{c}\tilde{\Delta}_{T}$ time series (dashed black) are displayed in (a). The amplitudes of the correlation functions are normalized. Probability density of ${}^{c}\tilde{A}$ (b) and ${}^{c}\tilde{\Delta}_{T}$ (c). The gray bars are the measures and the black line is the Gaussian fit.

objects can be simulated independently and freely associated. The properties of the resonant object (such as perceived material or shape) are included in the object part (as proposed by Gaver 1993) based on the synthesizer developed by Aramaki et al. (2006, 2009a).

In practice, the implementation process that associates the action with the object is based on a source-filter model, and the resulting sound is obtained by convolving the force (source) with the impulse response of a resonant object (filter bank). It is important to note that such a process based on a source-filter model does not limit the synthesis possibilities to the simulation of linear interaction phenomena. Indeed, nonlinear effects that are perceptually relevant can be taken into account in the source part. This has, for instance, been successfully accomplished for piano tone synthesis (Bensa, Jensen, and Kronland-Martinet 2004) and for synthesis of flute sounds (Ystad and Voinier 2001), but also for synthesis of nonlinear friction sounds (such as squeaking sounds, see Thoret et al. 2013). The implementation of the action and object parts will be described in the following sections.

Action Part Implementation

As previously presented, the force f that conveys the perceptual information about the interaction type (rolling, scratching, or rubbing) can be modeled as an impact series (Equation 2). The specific behaviors of the amplitude A^n and time interval Δ_T^n series seem to be an important perceptual cue associated with these interactions. We will experimentally characterize the time series ${}^{c}A = A - \mu_A$ or ${}^{c}\Delta_T = \Delta_T - \mu_{\Delta_T}$ (i.e., the centered versions of A and Δ_T) in order to, first, delineate a synthesis scheme for rolling sounds. Then we will describe how this model can be extended to simulations of rubbing and scratching sounds.

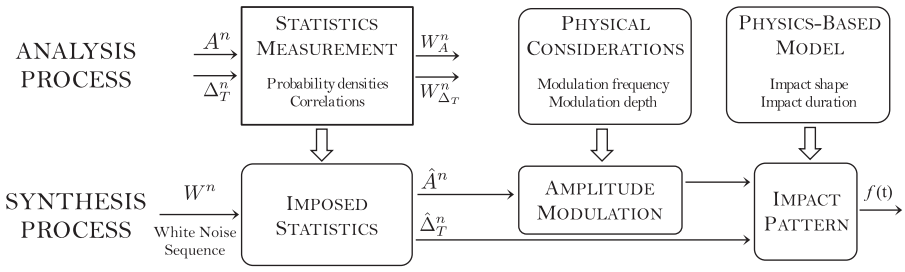
As pointed out in the section on the rolling invariants, A and Δ_T are strongly autocorrelated (as are ${}^{c}A$ or ${}^{c}\Delta_T$). We consider these two time series as autoregressive moving average processes. To characterize their behaviors, these series are "whitened" (as in Ninness, Wills, and Gibson 2005), and we experimentally noted that the whitening filters need have no more than one pole and one zero. Let X be one of the two processes ${}^{c}A$ or ${}^{c}\Delta_T$; we can then write in the z domain:

$$X(z) \approx H_X(z)\tilde{X}(z), \quad H_X(z) = \frac{1 + \sum_{i=1}^{p} b_i z^{-i}}{1 + \sum_{i=1}^{q} a_i z^{-i}}, \qquad (3)$$

where $\tilde{X}$ is the whitened version of X. As $\tilde{X}$ is white, we can properly estimate its probability density function and transform $\tilde{X}$ into W_X, which follows a uniform law, thanks to the inverse-transform sampling method (i.e., $W_X = F_X(\tilde{X})$, where F_X is the cumulative distribution function of the random variable $\tilde{X}$).

Figure 6a displays the autocorrelation of ${}^{c}\tilde{\Delta}_T$ (black line) and ${}^{c}\tilde{A}$ (gray line), the whitened version of ${}^{c}\Delta_T$ and ${}^{c}A$. The A and Δ_T time series were obtained from the force f computed with the physics-based model. As experimentally observed, this figure shows that a one-pole, one-zero whitening filter is well suited to whiten the time series. The probability densities of ${}^{c}\tilde{A}$ and ${}^{c}\tilde{\Delta}_T$ are plotted in Figure 6b and Figure 6c,

Figure 7. Schematic representation of the analysis–synthesis process.



respectively, showing that for the rolling force, they can be modeled as Gaussian processes. The analysis process of the impact series is schematically described in Figure 7.

In Figure 6a, $C_{(W_A^n, W_{\Delta_T}^n)}(k)$ the cross-correlation between the whitened process W_A and W_{Δ_T} is displayed. Given the long-term autocorrelation of ${}^cA$ and of ${}^c\Delta_T$, we assume that $W_{\Delta_T}^{n-1}$ has a poor influence on W_A^n (and, similarly, that W_A^{n-1} has a poor influence on $W_{\Delta_T}^n$), and that $C_{(W_A^n, W_{\Delta_T}^n)}(k)$ is proportional to $\delta(k)$, the unit impulse that equals 1 if $k = 0$ and equals 0 elsewhere. This led to the proposition of the synthesis scheme presented in Figure 7 where, according to the previous observations on $C_{(W_A^n, W_{\Delta_T}^n)}(k)$, we start from a single white and uniform process W to synthesize both $\hat{A}$ and $\hat{\Delta}_T$ (the hat designates for "estimated values," and no hat designates "measured values").

A sinusoidal modulation of the amplitude time series is further included in the synthesis process:

$$s(t) = [1 + m \sin(2\pi \nu_m t)], \tag{4}$$

with $m \in [0, 1]$ the modulation depth, controlling the perceived asymmetry of the rolling ball. The modulation frequency follows:

$$\nu_m \propto \frac{v}{R} \tag{5}$$

where v is the transversal velocity and R the ball radius (i.e., the modulation frequency increases with the rolling object's speed and decreases with its size.) As previously explained, this is done to emphasize the perceived velocity and size, as proposed in several studies (Hermes 1998; Houben 2002; Rath and Rocchesso 2005).

Finally, to simulate the whole force, we consider the "impact pattern" ϕ^n (cf. Equation 2) defined as a raised cosine function (van den Doel et al. 2001) :

$$\phi^n(t) = \begin{cases} \frac{1}{2}\left[1 + \cos\left(\frac{2\pi t}{t_0^n}\right)\right], & t \in \left[-\frac{t_0^n}{2}, \frac{t_0^n}{2}\right], \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

where t_0^n is the duration of the nth impact. As previously discussed, the impact duration varies with the impact amplitude, and we suggest the following relation:

$$t_0^n = \zeta \cdot (A^n)^{-\theta}, \tag{7}$$

where ζ is a constant that depends on the mass of the ball and on the stiffness k (see Equation 1), A^n is the n^{th} impact amplitude, and θ is a positive value that controls the strength of the dependence between the impact amplitude and the duration. This mapping was chosen based on several physical studies of impact sounds (Chaigne and Doutaut 1997; Avanzini and Rocchesso 2001).

Although this model was derived to synthesize rolling sounds, it is also well suited to reproduce rubbing and scratching sounds. As previously highlighted, the latter interaction forces are satisfactorily reproduced with stochastic processes such as white noise. Hence, by ignoring correlations between impacts, which in the present model corresponds to setting the (a_1, b_1) coefficients of the filters $H_A(.)$ and $H_{\Delta_T}(.)$ (Equation 3) to zero. In this case, the amplitude series A follows a Gaussian process and Δ_T follows an exponential distribution.

Object Part Implementation

The synthesis of impacted resonant objects was previously addressed by using an additive synthesis process (Aramaki and Kronland-Martinet 2006; Aramaki et al. 2006, 2009b). In this case, the frequencies of the oscillators corresponded to the eigenfrequencies of the resonant object. Noise could be added at the output of the oscillators to simulate the stochastic part of the impact. Then, the summed signal of noise and sinusoids was filtered into different frequency bands according to the Bark scale (Zwicker and Fastl 1990). A different time-varying envelope was applied to each band to take into account the frequency dependency of the damping. This synthesis process is effective for simulating single impacts, but not suitable when increasingly complex interactions are to be combined with the object. For instance, to synthesize bouncing objects, the signal of noise plus sinusoids has to be triggered precisely, as do the time-varying gains of the filter bank. This becomes computationally too expensive for more complex and continuous interactions such as rubbing, scratching, or rolling.

To overcome this limitation, a subtractive synthesis process is used to implement the object part. A resonant filter bank, whose frequencies and damping coefficients are tuned according to the eigenmodes and material properties of the resonant object, is used to simulate the object's impulse response. Each resonant filter's impulse response is an exponentially decaying sinusoid (Mathews and Smith 2003). With this approach, complex and continuous interactions can be fed into the resonant filter bank in a computationally efficient manner. A similar synthesis algorithm has already been proposed by van den Doel and Pai (2003).

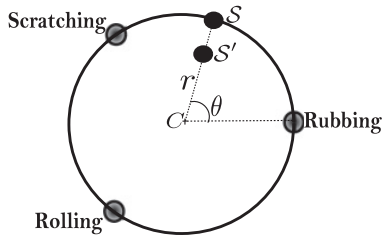
Intuitive Sound-Synthesis Control

In this section, we detail our control strategy for continuous navigation between rubbing, scratching, and rolling interactions. The intuitive control of the perceived material of the object part was suggested by Aramaki et al. (2011), and the interface allows the user to morph between different material categories (i.e., wood, metal, or glass) by moving a cursor in a two-dimensional space. The material categories are represented by anchors (prototypes) in the navigation space and correspond to specific sets of parameters. Thus, by interpolating spectral and damping parameters between the anchors according to the cursor's displacement, the synthesis process enables one to morph between materials and to create in-between (or hybrid) materials. Note that high-level control of the resonant object's shape is also available and allows the user to specify whether the object is one-dimensional (string-bar), two-dimensional (plate-membrane), or three-dimensional (shell).

Similarly to this material space for the control of objects, we present a navigation space for an intuitive control of the interaction type based on the definition of a "prototype" for each interaction. We will present here the navigation strategy between these prototypes, which are defined from the low-level parameters of the synthesis model as:

Impact model—two parameters control the impact duration t_0^n (Equation 7) in the chosen impact model (Equation 6): ζ and θ.
Probability density—the probability density is sampled as a set of discrete values, which are used to derive the cumulative distribution function F_X (the cumulative sum of the probability density), giving ${}^c\hat{\tilde{A}}$ and ${}^c\hat{\tilde{\Delta}}_T$ series. Two sets are defined, for the ${}^c\hat{\tilde{A}}$ and ${}^c\hat{\tilde{\Delta}}_T$ series respectively, and are written as $\mathbb{P}_A$ and $\mathbb{P}_{\Delta_T}$.
Whitening filters—as previously pointed out, one pole and one zero are sufficient for these filters. Each filter is described by a set of two coefficients (a_1, b_1), and each set is noted $\mathcal{C}_A$ and $\mathcal{C}_{\Delta_T}$, for $H_A(.)$ and $H_{\Delta_T}(.)$, respectively.
Offset coefficients—these correspond to the centered values μ_A and μ_{Δ_T} for $\hat{A}$ and $\hat{\Delta}_T$, respectively.
Amplitude modulation—the parameters are the modulation depth m and the frequency modulation ν_m (depending on the size and speed of the rolling object).
The global set of parameters is called: $\mathfrak{P} = \{\zeta, \theta, \mathbb{P}_A, \mathbb{P}_{\Delta_T}, \mathcal{C}_A, \mathcal{C}_{\Delta_T}, \mu_A, \mu_{\Delta_T}\}$.

Figure 8. Schematic control space of the interaction sound synthesizer.



Prototypes for Interaction Sounds

In this section, we will offer a set of parameters that can be used to calibrate "prototypes" for the three interaction sounds. Regarding the rubbing and scratching interactions, the precise parameter values have been documented in another paper (Conan et al. 2012). Regarding the rolling interaction, we have documented the parameter values, as well as the rolling controls such as size, velocity, and surface roughness, in a recent article published in *IEEE/ACM Transactions on Audio, Speech, and Language Processing* (Conan et al. 2014).

Rubbing Prototype

As shown by Conan et al. (2012) and noted in the section on sound invariants, small time intervals between impacts are perceptually associated with rubbing sounds (a maximum impact density implies a source signal that is white noise). For the rubbing sound prototype, we set the parameters to obtain a Gaussian white noise: the coefficients of $\mathcal{C}_A$ and $\mathcal{C}_{\Delta_T}$ were equal to zero, $\mu_A = \mu_{\Delta_T} = 0$, $\theta = 0$ (no dependency between impact amplitude and duration), $\mathbb{P}_{\Delta_T}$ was defined to get one impact at each sample, and $\mathbb{P}_A$ to follow a Gaussian distribution. The parameter ζ is set to get an impact duration of one sample and is used to control the perceived size of the object that rubs (longer impact durations, corresponding to low-pass filtering, evoke bigger objects). This set of parameters is called $\mathfrak{P}_{\text{rub}}$.

Scratching Prototype

The set of parameters for the scratching sound prototype is the same as for the rubbing prototype except for the Δ_T^n time series. As described previously, the scratching is associated with a low impact density (i.e., to high Δ_T^n values), and we propose that the Δ_T series follows an exponential distribution (defined empirically by Conan et al. 2012). This parameter set is called $\mathfrak{P}_{\text{scratch}}$.

Rolling Prototype

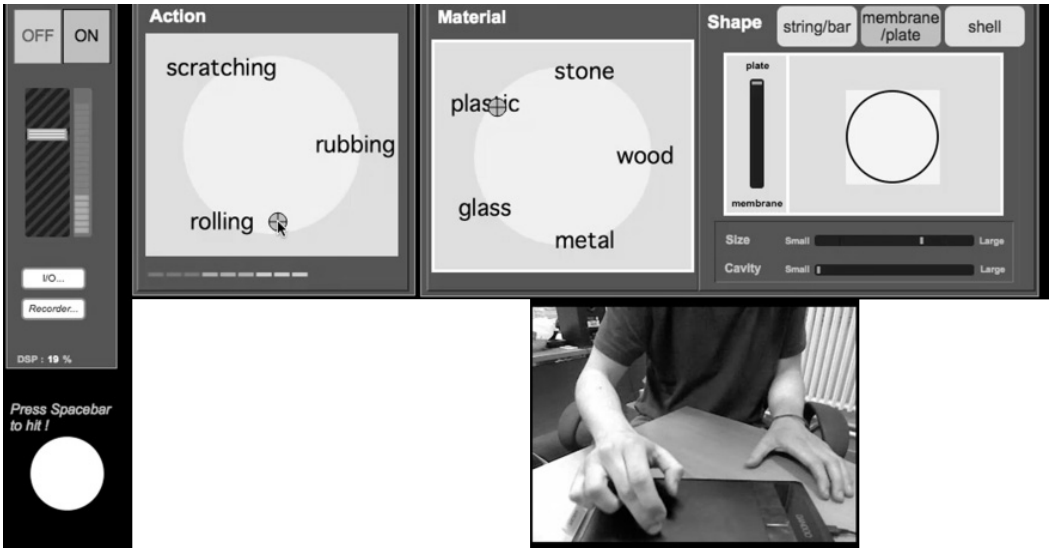
We set $\mathbb{P}_A$ and $\mathbb{P}_{\Delta_T}$ as Gaussian distributions. The values of the filter coefficients $\mathcal{C}_A$ and $\mathcal{C}_{\Delta_T}$, which are nonzero here, are based on the experimental analysis of the rolling force previously simulated. θ is set to 0.29, based on numerical simulations of the impact model (Equation 1). From a perceptual point of view, as ζ controls the contact duration, it is related to the size of the rolling ball. The filter coefficients $\mathcal{C}_A$ and $\mathcal{C}_{\Delta_T}$, as well as the probability densities $\mathbb{P}_A$ and $\mathbb{P}_{\Delta_T}$, are linked to surface parameters such as roughness. The perceived asymmetry lies within the interval [0, 1]. The modulation frequency is set to $\nu_m = 3V/S$, with $S \in]0, 1]$ the perceived ball size and $V \in [0, 1]$ the perceived ball velocity. This parameter set is called $\mathfrak{P}_{\text{roll}}$.

Navigation Strategy

The control strategy that we designed is inspired by the one described by Aramaki and colleagues (2011) to control the perceived material in an impact sound synthesizer. The three sound prototypes of perceived material (wood, metal, and glass, for which synthesis parameters were determined based on behavioral and electrophysiological experiments) are placed on the border of a disk that represents the so-called material space. The user can thereby navigate continuously between wood, metal, and glass by moving a cursor in this space, and the synthesis parameters are interpolated according to the distance to the three prototypes. A similar control space dedicated to interaction sounds is schematized in Figure 8.

The sound prototypes for rubbing, scratching, and rolling are placed as anchors on the circumference of a unit disk, at angles of 0, $2\pi/3$, and $4\pi/3$, respectively. Along this circumference, a sound $\mathcal{S}$, characterized by its angle θ, is defined by the

Figure 9. Synthesizer interface. One can see the user controlling the synthesizer by means of a graphical tablet that captures gesture velocity.



parameters $\mathfrak{P}_{\mathcal{S}}(\theta)$ as follows:

$$\mathfrak{P}_{\mathcal{S}}(\theta) = T(\theta)\mathfrak{P}_{\text{rub}} + T\left(\theta - \frac{2\pi}{3}\right)\mathfrak{P}_{\text{scratch}} + T\left(\theta - \frac{4\pi}{3}\right)\mathfrak{P}_{\text{roll}}. \quad (8)$$

The function $T(\theta)$ is defined as follows:

$$T(\theta) = \begin{cases} -\frac{3}{2\pi}\theta + 1, & \theta \in \left[0; \frac{2\pi}{3}\right[\\ 0, & \theta \in \left[\frac{2\pi}{3}; \frac{4\pi}{3}\right[\\ \frac{3}{2\pi}\theta - 2, & \theta \in \left[\frac{4\pi}{3}; 2\pi\right[\end{cases}, \quad (9)$$

Inside the disk, a sound $\mathcal{S}'$, characterized by both its angle θ and radius r, is defined by the parameters $\mathfrak{P}_{\mathcal{S}'}$:

$$\mathfrak{P}_{\mathcal{S}'}(\theta, r) = (1 - r)\mathfrak{P}_C + r\mathfrak{P}_{\mathcal{S}}(\theta), \quad (10)$$

where

$$\mathfrak{P}_C = \frac{1}{3}(\mathfrak{P}_{\text{rub}} + \mathfrak{P}_{\text{scratch}} + \mathfrak{P}_{\text{roll}}), \quad (11)$$

and $\mathfrak{P}_{\mathcal{S}}(\theta)$ is as defined in Equation 8.

In addition to the navigation in this "action space," the gesture is taken into account in the control strategy. Indeed, for such continuous interactions, the underlying gesture is a fundamental attribute that can be conveyed in the dynamics of the sound (Merer et al. 2013; Thoret et al. 2014). Following the synthesis process discussed by van den Doel, Kry, and Pai (2001), the resulting interaction force is low-pass filtered with a cutoff frequency that is directly related to the relative transversal velocity between the object that interacts (hand, plectrum, etc.) and the surface. When associated with a biological law, a specific calibration of the velocity profile enables the evocation of a human gesture (Thoret et al. 2014). The synthesizer interface is displayed in Figure 9. A demonstration video showing intuitive navigation in the action space and a gestural control on a graphic tablet is available online at www.lma.cnrs-mrs.fr/~kronland/CMJ2014 and is also at www.mitpressjournals.org/doi/suppl/10.1162/COMJ_a_00266/.

Conclusions

In this article, we described specific signal morphologies that are related to the auditory perception

of rubbing, scratching, and rolling interactions. Phenomenological considerations, physical modeling, and qualitative signal analysis were investigated, and we concluded that the interaction forces conveyed the relevant perceptual information about the type of interaction. A generic synthesis model aiming at reproducing these interaction forces (characterized by particular statistics of impact series) was designed. Then an intuitive control space that enables continuous transitions between these interactions was described.

Further studies will be done to expand this control space to other interactions such as nonlinear friction (squeaking, squealing, etc.; also see Thoret et al. 2013; Avanzini, Serafin, and Rocchesso 2005). The synthesis parameters associated with rubbing and scratching could be refined from the analysis of a large set of recorded sounds by using methods such as those discussed by Lagrange, Scavone, and Depalle (2010). The influence of the physical velocity profiles (rolling marble in a bowl, sliding object on an inclined plate, etc.) on the perceived interaction can also be studied, as already done by Thoret and co-workers (2014) on the evocation of human gestures by using specific velocity profiles. Such synthesis tools and morphing capabilities are of interest for motor rehabilitation purposes (Danna et al. 2013).

More interestingly, the proposed action–object framework is suitable for the creation of *sound metaphors*. This means that by freely combining objects with actions, unheard-of action–object combinations could be synthesized. With further experiments on new sound textures, the salience of the sound morphologies related to the evocations of the actions highlighted in this study can be accurately examined. For instance, is it possible to modify a given sound texture by means of the highlighted invariants, so that this texture evokes a rolling interaction? The determination of such transformational methods is useful in various domains related to musical as well as sonification applications. In particular, direct applications can be found in the current "MétaSon" project, in which specific sound-design issues are raised by the automobile industry. Because an increasing number of electric cars making very little noise will be used in the future, they should be equipped with sounds that clearly evoke a rolling object and that are recognizable as potentially dangerous objects for pedestrians. The model offered, together with the intuitive control strategy, constitutes a relevant tool for such investigations.

Acknowledgments

This work was funded by the French National Research Agency (ANR) under the MétaSon: Métaphores Sonores (Sound Metaphors) project (ANR-10-CORD-0003) in the Contenus Numériques et Interactions (CONTINT) 2010 framework (information online at metason.cnrs-mrs.fr/home.html). The authors Conan, Thoret, Aramaki, Gondre, Ystad, and Kronland-Martinet are with the Laboratoire de Mécanique et d'Acoustique (LMA), Centre National de la Recherche Scientifique (CNRS), Unité Propre de Recherche (UPR) 7051, Aix-Marseille Université, and École de Centrale Marseille. Derrien is with the Université de Toulon, LMA-CNRS-UPR 7051.

References

Akay, A. 2002. "Acoustics of Friction." *Journal of the Acoustical Society of America* 111(4):1525–1548.

Aramaki, M., and R. Kronland-Martinet. 2006. "Analysis-Synthesis of Impact Sounds by Real-Time Dynamic Filtering." *IEEE Transactions on Audio, Speech, and Language Processing* 14(2):695–705.

Aramaki, M., et al. 2006. "A Percussive Sound Synthesizer Based on Physical and Perceptual Attributes." *Computer Music Journal* 30(2):32–41.

Aramaki, M., et al. 2009a. "Thinking the Sounds: An Intuitive Control of an Impact Sound Synthesizer." In *Proceedings of the International Conference on Auditory Display*, pp. 119–124.

Aramaki, M., et al. 2009b. "Timbre Perception of Sounds from Impacted Materials: Behavioral, Electrophysiological and Acoustic Approaches." In S. Ystad, R. Kronland-Martinet, and K. Jensen, eds. *Computer Music Modeling and Retrieval: Genesis of Meaning in Sound and Music*. Berlin: Springer, pp. 1–17.

Aramaki, M., et al. 2011. "Controlling the Perceived Material in an Impact Sound Synthesizer." *IEEE*

Transactions on Audio, Speech, and Language Processing 19(2):301–314.
Avanzini, F., and D. Rocchesso. 2001. "Modeling Collision Sounds: Non-Linear Contact Force." In *Proceedings of the COST-G6 Conference on Digital Audio Effects*, pp. 61–66.
Avanzini, F., S. Serafin, and D. Rocchesso. 2005. "Interactive Simulation of Rigid Body Interaction with Friction-Induced Sound Generation." *IEEE Transactions on Speech and Audio Processing* 13(5):1073–1081.
Ben Abdelounis, H., et al. 2010. "An Experimental Study on Roughness Noise of Dry Rough Flat Surfaces." *Wear* 268(1):335–345.
Bensa, J., K. Jensen, and R. Kronland-Martinet. 2004. "A Hybrid Resynthesis Model for Hammer-String Interaction of Piano Tones." *EURASIP Journal on Applied Signal Processing* 2004(7):1021–1035.
Böttcher, N. 2013. "Current Problems and Future Possibilities of Procedural Audio in Computer Games." *Journal of Gaming and Virtual Worlds* 5(3):215–234.
Carello, C., K. Anderson, and A. Kunkler-Peck. 1998. "Perception of Object Length by Sound." *Psychological Science* 9(3):211–214.
Chaigne, A., and V. Doutaut. 1997. "Numerical Simulations of Xylophones, I: Time-Domain Modeling of the Vibrating Bars." *Journal of the Acoustical Society of America* 101(1):539–557.
Conan, S., et al. 2012. "Perceptual Differences between Sounds Produced by Different Continuous Interactions." In *Proceedings of the Acoustics 2012 Nantes Conference*, pp. 581–586.
Conan, S., et al. 2014. "A Synthesis Model with Intuitive Control Capabilities for Rolling Sounds." *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 22(8):1260–1273.
Danna, J., et al. 2013. "Handwriting Movement Sonification for the Rehabilitation of Dysgraphia." In *Proceedings of the International Symposium on Computer Music Multidisciplinary Research*, pp. 200–208.
Dubus, G., and R. Bresin. 2013. "A Systematic Review of Mapping Strategies for the Sonification of Physical Quantities." *PLOS ONE* 8(12):e82491.
Falcon, E., et al. 1998. "Behavior of One Inelastic Ball Bouncing Repeatedly Off the Ground." *The European Physical Journal B: Condensed Matter and Complex Systems* 3(1):45–57.
Farnell, A. 2010. *Designing Sound*. Cambridge, Massachusetts: MIT Press.
Gaver, W. 1993. "How Do We Hear in the World? Explorations in Ecological Acoustics." *Ecological Psychology* 5(4):285–313.
Gibson, J. J. 1966. *The Senses Considered as Perceptual Systems*. Boston: Houghton Mifflin.
Gibson, J. J. 1979. *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin.
Giordano, B., and S. McAdams. 2006. "Material Identification of Real Impact Sounds: Effects of Size Variation in Steel, Glass, Wood, and Plexiglass Plates." *Journal of the Acoustical Society of America* 119(2):1171–1181.
Gounaropoulos, A., and C. Johnson. 2006. "Synthesising Timbres and Timbre-Changes from Adjectives/Adverbs." In F. Rothlauf, et al., eds. *Applications of Evolutionary Computing*. Berlin: Springer, pp. 664–675.
Hermes, D. 1998. "Synthesis of the Sounds Produced by Rolling Balls." Internal IPO report no. 1226, IPO, Center for User-System Interaction, Eindhoven, The Netherlands.
Hoffman, M., and P. Cook. 2006. "Feature-Based Synthesis: Mapping Acoustic and Perceptual Features onto Synthesis Parameters." In *Proceedings of the International Computer Music Conference*, pp. 536–539.
Houben, M. 2002. "The Sound of Rolling Objects, Perception of Size and Speed." PhD dissertation, Technische Universiteteit, Eindhoven.
Hunt, K., and F. Crossley. 1975. "Coefficient of Restitution Interpreted as Damping in Vibroimpact." *Journal of Applied Mechanics* 42(2):440–445.
Johnson, K. L. 1987. *Contact Mechanics*. Cambridge: Cambridge University Press.
Klatzky, R., D. Pai, and E. Krotkov. 2000. "Perception of Material from Contact Sounds." *Presence: Teleoperators and Virtual Environments* 9(4):399–410.
Lagrange, M., G. Scavone, and P. Depalle. 2010. "Analysis/Synthesis of Sounds Generated by Sustained Contact between Rigid Objects." *IEEE Transactions on Audio, Speech, and Language Processing* 18(3):509–518.
Lakatos, S., S. McAdams, and R. Caussé. 1997. "The Representation of Auditory Source Characteristics: Simple Geometric Form." *Attention, Perception, and Psychophysics* 59(8):1180–1190.
Le Groux, S., and P. Verschure. 2008. "Perceptsynth: Mapping Perceptual Musical Features to Sound Synthesis Parameters." In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 125–128.
Lee, J., P. Depalle, and G. Scavone. 2010. "Analysis/Synthesis of Rolling Sounds Using a Source-Filter Approach." In *Proceedings of the International Conference on Digital Audio Effects*, pp. 483–487.

Lemaitre, G., and L. Heller. 2012. "Auditory Perception of Material Is Fragile While Action Is Strikingly Robust." *Journal of the Acoustical Society of America* 131(2):1337–1348.
Lloyd, D. B., N. Raghuvanshi, and N. K. Govindaraju. 2011. "Sound Synthesis for Impact Sounds in Video Games." In *Proceedings of the Symposium on Interactive 3D Graphics and Games*, pp. 55–62.
Mathews, M., and J. Smith. 2003. "Methods for Synthesizing Very High Q Parametrically Well Behaved Two Pole Filters." In *Proceedings of the Stockholm Musical Acoustics Conference*, pp. 405–408.
McAdams, S., and E. Bigand. 1993. *Thinking in Sound: The Cognitive Psychology of Human Audition*. Oxford: Oxford University Press.
Merer, A., et al. 2013. "Perceptual Characterization of Motion Evoked by Sounds for Synthesis Control Purposes." *ACM Transactions on Applied Perception* 10(1):1–24.
Michaels, C., and C. Carello. 1981. *Direct Perception*. Englewood Cliffs, New Jersey: Prentice-Hall.
Micoulaud-Franchi, J., et al. 2011. "Categorization and Timbre Perception of Environmental Sounds in Schizophrenia." *Psychiatry Research* 189(1):149–152.
Ninness, B., A. Wills, and S. Gibson. 2005. "The University of Newcastle Identification Toolbox (UNIT)." In *Proceedings of the World Congress of the International Federation of Automatic Control*. Available online at ww.sigpromu.org/reports/EE05010.pdf. Accessed June 2014.
Nordborg, A. 2002. "Wheel/Rail Noise Generation Due to Nonlinear Effects and Parametric Excitation." *Journal of the Acoustical Society of America* 111(4):1772–1781.
Papetti, S., F. Avanzini, and D. Rocchesso. 2011. "Numerical Methods for a Nonlinear Impact Model: A Comparative Study with Closed-Form Corrections." *IEEE Transactions on Audio, Speech, and Language Processing* 19(7):2146–2158.
Rath, M. 2004. "Interactive Realtime Sound Models for Human–Computer Interaction: A Sound Design Concept and Applications." PhD dissertation, University of Verona, Department of Computer Science.
Rath, M., and D. Rocchesso. 2004. "Informative Sonic Feedback for Continuous Human–Machine Interaction: Controlling a Sound Model of a Rolling Ball." *IEEE Multimedia Special on Interactive Sonification* 12(2):60–69.
Rath, M., and D. Rocchesso. 2005. "Continuous Sonic Feedback from a Rolling Ball." *IEEE Multimedia* 12(2):60–69.
Rocchesso, D., and F. Fontana. 2003. *The Sounding Object*. Florence: Mondo Estremo.
Rodger, M. W., W. R. Young, and C. M. Craig. 2013. "Synthesis of Walking Sounds for Alleviating Gait Disturbances in Parkinson's Disease." *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 22(3):543–548.
Stoelinga, C., and A. Chaigne. 2007. "Time-Domain Modeling and Simulation of Rolling Objects." *Acta Acustica United with Acustica* 93(2):290–304.
Thoret, E., et al. 2013. "Controlling a Non Linear Friction Model for Evocative Sound Synthesis Applications." In *Proceedings of the International Conference on Digital Audio Effects*, pp. 224–230.
Thoret, E., et al. 2014. "From Sound to Shape: Auditory Perception of Drawing Movements." *Journal of Experimental Psychology: Human Perception and Performance* 40(3):983–994.
Tucker, S., and G. Brown. 2002. "Investigating the Perception of the Size, Shape and Material of Damped and Free Vibrating Plates." Technical Report CS-02-10. University of Sheffield, Department of Computer Science.
van den Doel, K., P. Kry, and D. Pai. 2001. "Foleyautomatic: Physically-Based Sound Effects for Interactive Simulation and Animation." In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*, pp. 537–544.
van den Doel, K., and D. K. Pai. 2003. "Modal Synthesis for Vibrating Objects." In K. Greenebaum and R. Barzel, eds. *Audio Anecdotes III: Tools, Tips, and Techniques for Digital Audio*. Wellesley, Massachusetts: A K Peters, pp. 99–120.
Warren, W., and R. Verbrugge. 1984. "Auditory Perception of Breaking and Bouncing Events: A Case Study in Ecological Acoustics." *Journal of Experimental Psychology: Human Perception and Performance* 10(5):704–712.
Wildes, R. P., and W. A. Richards. 1988. "Recovering Material Properties from Sound." In W. A. Richards, ed. *Natural Computation*. Cambridge, Massachusetts: MIT Press, pp. 356–363.
Ystad, S., and T. Voinier. 2001. "A Virtually Real Flute." *Computer Music Journal* 25(2):13–24.
Zwicker, E., and H. Fastl. 1990. *Psychoacoustics: Facts and Models.* Berlin: Springer.

Journal of Experimental Psychology:
Human Perception and Performance
2014, Vol. 40, No. 3, 983–994

© 2014 American Psychological Association
0096-1523/14/$12.00 DOI: 10.1037/a0035441

From Sound to Shape: Auditory Perception of Drawing Movements

Etienne Thoret, Mitsuko Aramaki, and Richard Kronland-Martinet
Laboratoire de Mécanique et d'Acoustique UPR7051 CNRS, Marseille, France

Jean-Luc Velay
Laboratoire de Neurosciences Cognitives UMR7291 CNRS, Marseille, France

Sølvi Ystad
Laboratoire de Mécanique et d'Acoustique UPR7051 CNRS, Marseille, France

This study investigates the human ability to perceive biological movements through friction sounds produced by drawings and, furthermore, the ability to recover drawn shapes from the friction sounds generated. In a first experiment, friction sounds, real-time synthesized and modulated by the velocity profile of the drawing gesture, revealed that subjects associated a biological movement to those sounds whose timbre variations were generated by velocity profiles following the 1/3 power law. This finding demonstrates that sounds can adequately inform about human movements if their acoustic characteristics are in accordance with the kinematic rule governing actual movements. Further investigations of our ability to recognize drawn shapes were carried out in 2 association tasks in which both recorded and synthesized sounds had to be associated to both distinct and similar visual shapes. Results revealed that, for both synthesized and recorded sounds, subjects made correct associations for distinct shapes, although some confusion was observed for similar shapes. The comparisons made between recorded and synthesized sounds lead to conclude that the timbre variations induced by the velocity profile enabled the shape recognition. The results are discussed in the context of the ecological and ideomotor frameworks.

Keywords: biological motion, action-perception, friction sound synthesis

Supplemental materials: http://dx.doi.org/10.1037/a0035441.supp

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

The perception of movement induced by sounds is a widely investigated research topic involving researchers from diverse areas, ranging from physics to cognitive neuroscience. However, compared with other aspects of movement investigation, one aspect seems to have been accorded little attention, namely the auditory perception of biological movement. This is the subject of the present investigation. In the following, we give an overview of former studies and theoretical frameworks that have dealt with the auditory perception of acoustical events and the perception and production of biological movements.

From an acoustical point of view, the auditory perception of movement induced by one or several sound sources was initially addressed by investigating the auditory consequence of an actual displacement of a physical source. It is well known for instance, that the perception of a passing source is related to the physical phenomena occurring during sound propagation. Hence, a sound emanating from a source at a distance from the listener is less intense, more band-limited and more reverberant than a sound from a nearby source. As the source approaches the listener, its intensity and bandwidth increase in parallel, and its level of reverberation decreases. In addition, frequency shifts, due to the Doppler effect, occur when there is a change in the speed of the moving source relative to the listener. Therefore, the overall variation of the generated sound is determined by combining variations in intensity, frequency, bandwidth, and reverberation. Such sound can be reproduced satisfactorily under monophonic playback conditions, which implies that specific sound morphologies, mainly related to the timbre, can evoke movement.

An audio researcher, a sound engineer, or a musician is usually required to reproduce such types of effects by manipulating the intrinsic characteristics of the sound. These characteristics were widely exploited for sound modeling purposes and music composition. Therefore, it is reasonable to consider that evoked movements not only concern the physical displacement of a source but

This article was published Online First January 20, 2014.

Etienne Thoret, Laboratoire de Mécanique et d'Acoustique, CNRS, UPR 7051, Aix-Marseille University, Centrale Marseille, Marseille; Mitsuko Aramaki, Laboratoire de Mécanique et d'Acoustique, CNRS, UPR 7051, Aix-Marseille University, Centrale Marseille, Marseille; Richard Kronland-Martinet: Laboratoire de Mécanique et d'Acoustique, CNRS, UPR 7051, Aix-Marseille University, Centrale Marseille, Marseille; Jean-Luc Velay, Laboratoire de Neurosciences Cognitives, CNRS, UMR 7291, Aix-Marseille University, Marseille; Sølvi Ystad, Laboratoire de Mécanique et d'Acoustique, CNRS, UPR 7051, Aix-Marseille University, Centrale Marseille, Marseille.

This work was funded by the French National Research Agency (ANR) under the MetaSon: Métaphores Sonores (Sound Metaphors) project (ANR-10-CORD-0003) in the CONTINT 2010 framework. We thank the reviewers for their helpful comments. The authors are also grateful to Charles Gondre for his precious help in developing of the listening test interfaces, and to Lionel Bringoux for his powerful remarks.

Correspondence concerning this article should be addressed to Etienne Thoret, Laboratoire de Mécanique et d′Acoustique, CNRS, UPR 7051, Aix-Marseille University, Centrale Marseille, 31, Chemin Joseph Aiguier, F-13402 Marseille cedex 20. E-mail: thoret@lma.cnrs-mrs.fr

may also refer to more metaphoric notions of movement, like musical movements, for instance. Indeed, music analyses have led to semiotic descriptions of perceived movements in musical pieces (Frémiot et al., 1996). Other studies have focused on the nature the relationships between music and motion in general (Honing, 2003; Johnson & Larson, 2003). To investigate sound attributes related to the general concept of motion evoked by sounds, Merer, Ystad, Kronland-Martinet, and Aramaki (2008) conducted a free categorization task that used monophonic abstract sounds, that is to say, sounds whose source was not readily identifiable. On the basis of this, they identified the main movement categories such as "Rotate," "Pass by," or "Fall down," which were associated to different acoustic descriptors, such as the ratio of frequency modulation or the amplitude modulation for instance. Merer et al. further addressed the perceptual characterization of these evoked motions by studying the drawings produced by a group of subjects using a purpose-made graphical user interface while listening to sounds (Merer, Aramaki, Ystad, & Kronland-Martinet, 2013). Based on an analysis of the drawings, some perceptually relevant variables accounting satisfactorily for the motion perceived in the sounds were identified.

From a theoretical perspective, according to the ecological theory of perception, the acoustic properties that carry the information for sound source identification are known as invariants. Originally introduced for vision, Gibson (1966) defined the invariants as the properties of the environment that don't vary, and thus, reveal a structure in a sensorial flow enabling perception and action. Following this ecological perspective, Gaver (1993b) proposed that perceiving a sound event is more than just a pattern matching with memorized representations and that sound inherently conveys consistent information about the physical world through invariant acoustic features, such as a temporal pattern or a spectral relationship, contained in the acoustic flow. More precisely, the information that specifies the nature of the sound source is known as a *structural invariant* and the information specifying the type of change or the action involved is referred as the *transformational invariant* (McAdams, 1993).

Concerning structural invariants, many studies have investigated the relations between physical characteristics of objects and the perception of the sound produced when they are impacted (e.g., Giordano, Rocchesso, & McAdams, 2010; Grassi, 2005; Grassi, Pastore, & Lemaitre, 2013; Kunkler-Peck & Turvey, 2000; McAdams, Chaigne, & Roussarie, 2004). Some of these studies revealed certain physical characteristics linked to dispersion and dissipation that are important for the recognition of a sound source. It has been shown, for instance, that sounds contain sufficient information to enable one to discriminate the material of impacted objects (Aramaki, Besson, Kronland-Martinet, & Ystad, 2011; Klatzky, Pai, & Krotkov, 2000; Wildes & Richards, 1988), and, to a certain extent, to recognize their shapes (Carello, Anderson, & Kunkler-Peck, 1998; Lakatos, McAdams, & Caussé, 1997). In particular, Giordano and McAdams (2006) evaluated the effect of the size of an object on the perception of the material for an impact sound and identified robust acoustical descriptors that explain material identification.

Concerning transformational invariants (Gaver, 1993a), they are related to the actions carried out on a given object. For instance, Warren and Verbrugge (1984) showed that based on the rhythm of a series of impacts contained in a sound, it is possible to predict if a glass had broken or bounced. Li, Logan, and Pastore (1991) revealed that one is able to recognize the gender of a person walking merely by listening to the footstep sounds produced, and more precisely that this was due to the differences in spectral peaks and the contribution of high frequency components for the feminine gender. Repp (1987) concluded that the sound of two hands clapping was sufficient to imagine the spatial conformation of the two hands.

For its part, the structural invariant of an object can be recognized even if it is associated with different transformational invariants. Hence, a ball is still recognized as such even if it is submitted to different actions like bouncing or rolling. Similarly, several studies demonstrated that the sound produced by a rolling ball could reveal its perceived size or velocity (Houben, Kohlrausch, & Hermes, 2004, 2005). Note that a study by Lemaitre and Heller (2012) highlighted that our auditory system is better tuned to recognize an object's action than its material based on the sounds produced by different interactions: rubbing, rolling, bouncing, or impacting.

In the ecological perspective, the concept of invariant is formalized with the notion of affordances, which can be defined as the ability of an object, here a sound, to evoke its use. In the case of sounds, the extracted invariants afford the potential actions that enable recognition and categorization of an action. Therefore, links between perceptual abilities and actions, here concerning auditory perception, are primordial (Castiello, Giordano, Begliomini, Ansuini, & Grassi, 2010).

Although fully compatible with the ecological approach described above, these abilities were also regarded as arising from cognitive processes involving the concept of representation. The case of speech perception is an appropriate example of such a sensorimotor coupling between the auditory perception of an event, and the motor representation that is inferred. Indeed, motor theory of speech perception suggests that we do not perceive sounds exclusively as auditory information, but that we perceive it as potential intended phonetic gestures (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Liberman & Mattilngly, 1985). It has been shown that we have learned to bind the sounds produced by objects and modulated by actions with all their other perceptual properties in order to create a unified percept (Hommel, 2004). Recent neuroscientific research has shed light on brain mechanisms underpinning such percepts that might arise from multimodal neurons coding of both movements and their sensory consequences. Especially, it has been shown that mirror neurons in the monkey ventral premotor cortex discharge when the animal performs a specific action, but also when it hears the corresponding action-related sound without seeing the action in question (Kohler et al., 2002). However, the movements have to respect the relevant rules of production, otherwise the generated sound can lead to a misinterpretation of the action. More recently, Young, Rodger, and Craig (2013) discussed auditory-motor relations involved in the real time reproduction of walking sounds in the ideomotor framework. This perspective considers cognitive representations as a structural coupling between perceptions and actions (Hommel, Müsseler, Aschersleben, & Prinz, 2001; Prinz, 1997). In Young et al. (2013), they proposed that the ability to synchronize our walking with a walking sound is made possible thanks to a common neural representation of the perceived and the generated action with regards to recent models (Cisek & Kalaska, 2010). As such, this novel approach tends to reconcile direct and indirect perspectives to perception and action (Norman, 2002).

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

In the present work, we focus on the auditory perception of a particular type of movement belonging to a category of biological movement, that is, drawing movements, which are, essentially, specific to humans. Because drawings generate a visual trace, but no dominant sound, these movements are far less connected to audition than they are to vision. Indeed, the quality of drawing is judged, not by its sound, but by its visual trace. However, if we listen carefully to the sound produced during drawing, we can hear the pen friction against the paper, especially when the surface is rough. However, given that it is unusual to pay attention to such sounds, it seems, a priory, very difficult, to say the least, to try to infer from these sounds what is drawn or written. Yet, these sounds are more than just noise. Due to the friction between the pen tip and the paper asperities, the kinematics of the drawing movement produce timbre variations in the sound that may, to a certain extent, enable the recognition of a specific movement produced by the writer. In particular, we wanted to find out if it would be possible to infer by ear what is drawn based, only, on the friction sounds generated. Our aim was thus to identify acoustic cues that reflect the movements underlying the drawing action and to ascertain the extent to which this information allows us to infer the characteristics of the drawn shapes.

These issues were investigated in three experiments. As a first step, we verified whether subjects were able to associate timbre variations in a friction sound to kinematic variations produced by the gesture during the drawing process. To this end, we began by consulting a number of seminal studies from Viviani and colleagues who carried out extensive investigations of the production and perception of biological movements in both the visual and sensorimotor modalities and, particularly, the relationship between graphical movements and geometric shapes. They highlighted the link between the velocity of drawing movements and the curvature of the drawn shape (Viviani & Terzuolo, 1982) and proposed a power law relation between the angular velocity of the pen and the curvature of the drawn trajectory (Lacquaniti, Terzuolo, & Viviani, 1983; Viviani & Flash, 1995; Viviani & McCollum, 1983).

In our first experiment (Experiment 1), we adapted a protocol of Viviani and Stucchi (1992) to the auditory perception of friction sounds. The subjects were asked to manipulate the power law modulating the friction sound (simulating the sounds produced by the pen) so as to imitate as closely as possible a sound evoking a natural and fluid drawing movement. This experiment required the use of real-time generated, synthesized stimuli. Our first hypothesis was, therefore, that timbre variations of a natural and fluid gesture should be recognizable through the exponent value of the power law. And thus, such timbre variations could be considered as an auditory transformational invariant enabling the recognition of some human gestures. As a second step, we speculated that, if a friction sound can indeed evoke a gesture, one might be able to visualize, to a certain extent, the geometric shape that has been drawn on the basis of the friction sound only. We thus asked subjects to associate friction sounds to given visual shapes in a series of experiments (Experiments 2 and 3); both recorded and synthesized sounds were evaluated. Using synthesized stimuli made it possible to focus on a single gesture parameter, namely the velocity profile, and to determine the extent to which this parameter is relevant for a sound/shape association task. In Experiment 2, the shapes were assumed to be easily distinguishable, both from a perceptual point of view and from the kinematics of their underlying drawing movements. In Experiment 3, we included sounds and shapes that were assumed to be perceptually more similar to each other to assess if similar velocity profiles imply lower recognition rates.

The results of the experiments will be discussed from an ecological perspective to determine whether the velocity profile can be considered as a relevant transformational invariant of drawing movement. In addition, the results will be discussed in light of the ideomotor framework proposed by Young et al. (2013), in particular in terms of representations as a structural coupling between the sensorial flow and its processing. Before presenting these three experiments, we will describe in the following section, the general principles of the synthesis of friction sounds.

Synthesis of Friction Sounds

Synthesis is an appropriate tool for investigating the perception of the underlying gesture evoked by friction sounds. Gaver (1993b) and Van den Doel, Kry, and Pai (2001), for example, proposed a simple, physically informed model for generating synthetic friction sounds from given velocity and pressure profiles. This model simulates the physical sound source resulting from successive impacts of a plectrum on the asperities of a surface, in the context of our study the movement of a pencil on paper. The surface roughness is modeled by a noise reflecting the heights of the surface asperities. A common model for such noise is the fractal model whose spectrum is: $S(f) = 1/f^b$, where b can range between 0 and 2 and f is the frequency (Van den Doel et al., 2001). If $b = 0$, the noise is white and corresponds to a rough surface, and as the value of b increases, the surface smoothens. In the following, we set $b = 0$. The sound was then generated either (a) by reading the noise (stored in a wavetable) which velocity is linked to the velocity profile of the pencil rubbing against the paper (and therefore to the velocity of the gesture); or (b) by lowpass filtering this noise with a cut-off frequency that varies according to the velocity profile of the pencil. This latter method is computationally more efficient. The mapping between the velocity of the pen (in $cm.s^{-1}$) and the synthesis model is arbitrary.

In addition to the velocity, other gesture parameters such as the pressure or the angle of the pencil could be considered in a more sophisticated model. For instance, the pressure of the pencil could be related to the intensity of the sound. However, because the focus of the current study is the velocity profile, the pressure is kept constant throughout the present studies. Thus, only the velocity profile of the gesture was used to control the production of the friction sound.

Experiment 1

As a first approach to investigating our capacity to recognize specific gestures through sounds, we based on a protocol of Viviani and Stucchi (1992) on the perception of visual biological motion. They asked subjects to adjust the velocity of a visual dot until they perceived its displacement as being uniform (i.e., a perceived constant speed). The movement of the dot was constrained by the power law expressed as:

$$v_t = KC^{-\beta} \quad (1)$$

where v_t is the tangential velocity, C the radius of the trajectory curvature and K, the velocity gain factor linked to the overall movement speed (Viviani & Terzuolo, 1982; Lacquaniti, Terzuolo,

& Viviani, 1983; Viviani & McCollum, 1983; Viviani & Flash, 1995). During the task, subjects were unaware that they were adjusting the exponent β of the power law. Results revealed that, to perceive a uniform visual displacement, the subjects adjusted the exponent by an average value of one third. This value corresponded to the actual velocity profile of the physical movement used during the drawing production. Conversely, when the exponent was null (i.e., constant speed regardless of the curvature, corresponding to a uniform physical movement), the movement of the dot was perceived as accelerating in the curved parts of the trajectory. In the following, we will refer to Equation (1) with $\beta = 1/3$ and call this relation the 1/3 power law. Note that this power law optimizes motion smoothness by minimizing the rate of change of acceleration (the jerk), as was previously shown by numerical simulation and mathematical analysis (Richardson & Flash, 2002; Viviani & Flash, 1995).

In the present study, we designed an experiment similar to that of Viviani and Stucchi (1992) in which subjects had to implicitly act on the exponent value of the power law within the auditory modality. In particular, the moving visual spot that described the trajectory, which was used in Viviani and Stucchi's study, was, here, replaced by a synthetic friction sound modulated according to the velocity profile. Although Viviani and colleagues aimed at investigating the perceptual relationship between the two visual variables, that is, the kinematics and the curvature, it should be noted that our goal was to investigate whether the manipulation of one variable related to the gesture's kinematics (determined by the velocity profile) allowed the evocation of a natural and fluid gesture. In particular, we assumed that the specific sound variations caused by the 1/3 exponent are recognizable through the auditory modality. If our hypotheses proved correct, this would imply that the timbre variations of the sound convey perceptual information about the physical movement, that is, the pen accelerating over the straight sections of the traced strokes and slowing down in the most curved sections.

Method

Subjects. Twenty participants took part in this experiment, 3 women and 17 men. Their average age was 29.42 years (SD = 12.54). Before participating in this experiment, none of the subjects were familiar with the topic being investigated.

Stimuli. Friction sounds were synthesized using the previously described friction model. To avoid evoking specific shapes, we considered friction sounds associated with pseudorandom trajectories generated from the trajectories of a moving point (x(t), y(t)), defined on the basis of the following parametric functions:

$$\begin{cases} x(t) = A_0 \sum_{k=0}^{3} a_k \sin(\omega_{x,k} t) \\ y(t) = B_0 \sum_{k=0}^{3} b_k \sin(\omega_{y,k} t) \end{cases} \quad (2)$$

A new set of parameter values for the above equations was randomly computed every 15 s (arbitrary choice), with the exception of the constant values $A_0 = 7$, $B_0 = 5$, $a_0 = 1$, and $b_0 = 1$. In particular, the values of $\omega_{x,k}$ and $\omega_{y,k}$ were randomized between 0 and 0.6 Hz and a_k and b_k were randomized between 0.5 and 0.9. Hence, the movement of the point was predictable for only 15 s. The velocity profile corresponding to this pseudorandom trajectory was then computed from the power law expressed in Equation (1), where $K = 10\ \mathrm{m.s}^{-1}$ and the curvature is defined by the following expression:

$$C(t) = \frac{|\dot{x}\ddot{y} - \dot{x}\ddot{y}|}{(\dot{x}^2 + \dot{y}^2)^{\frac{3}{2}}}$$

where x and y were defined in (2), the dot and double dot represent the first and the second time derivative, respectively. The corresponding friction sound was then synthesized in real-time according to the exponent value of the power law, which was modified directly by the subjects during the experiment. Examples of stimuli for four exponent values ($\beta = 0$; $\beta = 0.33$; $\beta = 0.7$; $\beta = 0.9$) are available in supplemental material online.

Task and procedure. Participants were seated in front of a computer screen in a quiet room. Sounds were presented through Sennheiser HD-650 headphones. The experiment began with a 2-trial training session, followed by a 6-trial session constituting the formal test. For each trial, a pseudorandom trajectory was computed, and the corresponding friction sound was synthesized in real-time. The trajectory varied across trials and subjects. The trajectory was not displayed to the subjects and only the sound was presented continuously during the trial.

Subjects were asked to imagine that the sound they heard was produced by someone drawing a random shape (such as a scribble) on a rough surface. They were then asked to modify the sound using two assigned buttons ("<" and ">" presented on the computer screen) until they arrived at a sound that they judged the most "natural" and "fluid," according to a human gesture. The graphical interface was designed with the real time software Max/MSP.[1] The subjects were unaware of how their interaction with the buttons modified the sound and they also did not know that they were actually adjusting the exponent β of the power law ("<" and ">" buttons corresponded to decreasing and increasing the β values, respectively). Subjects were advised to take their time when listening to the sound and to explore the full range of values with the buttons during the adjustment process. The exponent values ranged between 0 and 1.0816 in steps of 0.0416. The initial exponent values were randomized at each trial. No time constraint was imposed. For each subject, one exponent value for each of the six trials was collected.

Results

The exponent values were averaged across trials, first for each subject, and then across subjects. An average value of 0.361 (SD = .084) was found. A comparison of this mean to one third was then performed by means of a one-sample two-tailed t-test. For statistical analyses, effects were considered significant if the p value was equal to or less than .05.

No significant differences between the mean values of the distributions were found, $t(19) = 1.53$; $p = .14$; $d = .329$. An analysis of the results according to the initial values of the exponent was performed to evaluate a possible ascending or descending effect on the subjects' performance. A classical effect of ascending

[1] http://cycling74.com

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

and descending threshold was observed and a significant correlation between the initial and the final values was found ($r = .34$; $p < .05$). As proposed in Carlyon et al. (2010), we performed a complementary analysis to confirm that exponent values had actually been adjusted to values close to one third. Two groups of final exponents were considered: trials in which the initial values were greater than one third and those in which they were lower than one third. Out of all the trials, 70 initial values were higher than one third and 50 were lower. For ascending thresholds, the mean value of the final exponent was .31 (95% CI [0.28; 0.34]); for the descending thresholds, the mean value of the exponent was .39 (95% CI [0.36; 0.43]). Moreover, we compared the standard deviations of the initial and final values with one-tailed two-sample t tests to check whether subjects adjusted the final exponents to a given value, so reducing their dispersion. The tests revealed that the final standard deviations were significantly lower than the initial ones, $t(19) = -7.86$; $p < .05$, confirming that the subjects actually carried out an adjustment.

Discussion

In this experiment, the subjects had to act on the sound they heard to render the evoked human gesture as "natural" and as "fluid" as possible. They were told that the sound was produced by someone drawing a shape with no specific geometry on a sheet of paper using a pen. The sound was generated by synthesis from the velocity profile expressed in Equation (1) and could be adjusted according to a hidden variable corresponding to the exponent β of the power law. Indeed, it is worth noting that the subjects were unaware of the acoustical characteristics of the sound they were adjusting with the control buttons.

The main finding of this first experiment is that, when asked to adjust a sound to evoke a "natural" and "fluid" graphical movement, subjects adjusted the timbre variations so that the velocity profile matches the 1/3 power law. In other words, the timbre of the sound of a moving pen appears to vary in accordance with the kinematic rule governing real graphical movements (Viviani & Terzuolo, 1982). This result reveals that we are able to imagine a human natural gesture, that is, biological movements, from the timbre variations of the sound in accordance with previous findings linked to the recognition of acoustic events through specific acoustical patterns, so-called transformational invariants. In other words, if one adopts a sensorimotor perspective, our perceptual processes are shaped by our motor competencies making us able to recognize such events and to interact with them (Viviani, 2002).

In this first experiment, subjects knew that the evoked movement corresponded to scribbles, and they were not required to make inferences regarding the actual shapes drawn. Because the 1/3 power law linking the velocity of the gesture and the curvature of the drawn trajectory was relevant for the evocation of natural human gestures through sounds, it may be possible, to a certain extent, to evoke a given trajectory from the friction sound and to make associations between the drawn shapes and friction sounds. Indeed, the association process may be based on the fact that higher velocities are due to smaller curvatures and conversely. Although only one variable was manipulated for the synthetic sound, that is, the velocity profile, the implicit knowledge of the 1/3 power law should enable an association between a visual depiction of the drawn shape and the friction sound. This was the focus of our second and third experiments.

Experiment 2

In drawing, a shape can be described both by its visual geometry and, in motor terms, by the movement required to trace it. Freyd (1983a; 1983b) first demonstrated that a motor representation is intrinsically linked to the perception of the visual shape. In particular, she found that readers use motor knowledge when decoding static, hand-written material. Furthermore, a close functional relation between the visual shape of a character and the corresponding graphic movement has been established: It was found that looking at graphic shapes activates cortical motor processes if the subjects already know how to draw these shapes (James & Gauthier, 2006; Longcamp, Anton, Roth, & Velay, 2003; Longcamp, Tanskanen, & Hari, 2006). Longcamp et al. (2008) assumed that the specific movements used to write a novel graphic shape are memorized and are, furthermore, involved in its subsequent visual recognition. In line with these results, Viviani and Stucchi (1989) established that a perceived shape, described by the displacement of a punctual spot, can be accurately deduced from the kinematics of the moving spot. Furthermore, they highlighted cases in which the perception of a moving spot could interfere with the perception of the resulting shape. For instance, when a point is moving along an elliptic trajectory with a small eccentricity at a constant speed, that is, without respecting the 1/3 power law, the perceived ellipse is modified and the resulting shape was assimilated as a circle.

By contrast, no sound is naturally associated with a visual shape because a shape is an abstract object and therefore cannot be readily associated with a physical sound source. In Experiment 1, we confirmed the relevance of the 1/3 power law in making the sound coincide naturally with a natural and fluid human gesture. Because this law links the velocity of the movement and the curvature of the drawn shape, we can assume that the pen trajectory was processed implicitly in the previous task.

In this second experiment, our aim was to investigate further whether friction sounds can also inform about the drawn shapes. We asked subjects to associate a given friction sound to a static visual shape they imagined was drawn during the sound production. Subjects had to choose between four visual shapes corresponding to the actual drawn ones. The rationale for this association task was to limit the number of possible shapes to be identified. Based on the results of Experiment 1, we hypothesized that the relevant information required by the subjects to associate the sounds with the shapes is contained in the velocity profile, and that, according to the 1/3 power law, we are able, to a certain extent, to recover geometric information about the drawn shapes from sounds. On this basis, we began by recording the natural sounds produced by a writer for different elementary drawn shapes. Indeed real friction sounds contain fine modulations that may vary, for example, according to the pen angle with respect to the paper. In addition, for downstrokes and upstrokes, variations of this angle cause variations in the pen pressure. To test our hypothesis and to investigate the relevance of the velocity profile alone, we also considered synthetic friction sounds that depended solely on the velocity profiles collected from a writer.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Method

Subjects. Twenty participants took part in the experiment: 9 women and 11 men. The average age was 30.65 years (SD = 13.11). None of the subjects were familiar with the topic of the study before the test. Eight subjects had also participated in Experiment 1.

Stimuli.

Static shapes. Preliminary informal tests revealed that distinguishing sounds corresponding to shapes with cusps from those without cusps was perceptually easy because cusps gave rise to a discontinuity in the sound produced, that is, a stop, that was highly perceptible. Based on these observations, we assembled a corpus of four shapes, two of which had no cusps (circle, ellipse) and two of which had cusps (arches, line). Shapes without cusps included a circle and an ellipse that differed by their eccentricity. Those with cusps were arches and line that differed by their cusp positions: the arches cusps were located periodically along the paper, and the line cusps were located at the same positions, that is, at the extremities. Figure 1 presents the four shapes.

Recorded sounds. Recording sessions took place in a quiet recording studio. A member of the staff drew the four shapes as fluidly as possible on a Wacom graphic tablet–Intuos 3. Excepted for the arches, he was asked not to lift the pen from the tablet for 25 s to make sure that several periods were drawn for each shape; for the arches he was asked to start drawing from the left side of the sheet until he reached its end, and then to start again at the initial position. The movements of the writer were recorded on the tablet at a sampling frequency of 200 Hz and with a spatial precision of 5.10^{-3} mm.

In addition, monophonic recordings of the sounds produced during the drawing sessions were made at a sampling frequency of 44100 Hz with a cardioid Neumann-KM84i microphone positioned about 30 cm above the tablet. Sequences of these recordings were selected for the experiment based on a geometrical and temporal examination of the writer's performances. In particular, we selected sequences of similar durations (about 5 seconds) during which the writer executed the shapes in a regular manner for a given number of periods. For ellipses, arches, and lines, the selected sequences corresponded to four periods. For circles, only two periods were considered because the mean duration was about twice that of the three other shapes. Table 1 summarizes the characteristics of the performances chosen as stimuli for the experiment. The segmentation was performed with a windowing function of 10 ms to avoid clicks at the beginning (fade in) and at the end (fade out) of each sound sequence. All the sounds were normalized at −3 dB. The velocity profiles corresponding to the selected sequences were computed from the tablet data for the four shapes. Figure 2 presents one period of the velocity profile of each shape. They are normalized in amplitude and low-pass filtered at 10 Hz. The zero-crossings corresponding to the cusps are marked

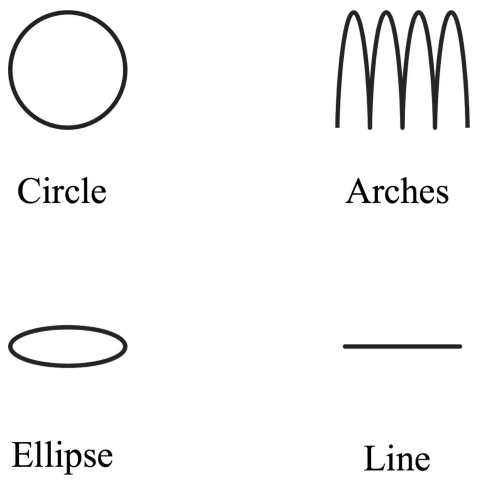


Figure 1. The four shapes used in Experiment 2. For arches, four periods are presented.

Table 1
Geometrical and Temporal Characteristics of the Performances Chosen for the Stimuli of the Experiment 2

Shape	Length (cm)	Duration (s)
Circle	62.5	5.2
Ellipse	89.32	5.8
Arches	87.11	5.1
Line	88.7	5.2

Note. The circle length, which corresponds to the drawn length used to synthesize stimuli, appears to be shorter than the three other shapes. However, the duration of the associated sound is almost the same as the three other ones because the mean velocity is lower for this shape. The recorded and synthesized stimuli of Experiment 2 are available in supplementary material online.

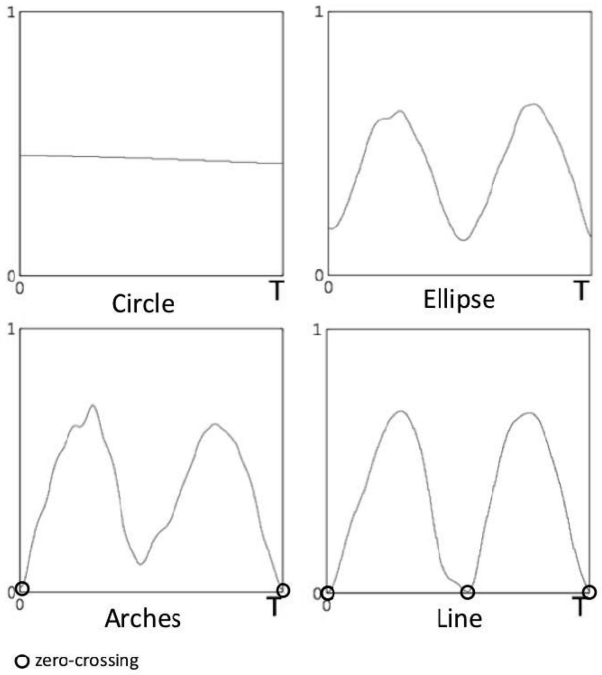


Figure 2. One period of the velocity profiles of the four shapes used in Experiment 2. The zero-crossings are marked with black circles. The velocity profiles are low pass filtered at 10 Hz and normalized in amplitude.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

with black circles. It is noticeable that the shapes have very different velocity profiles, which led us to hypothesize that they should be distinguishable from a perceptual point of view.

Synthesized sounds. The friction sounds were synthesized using the same friction model as in Experiment 1. They were generated from the velocity profiles collected from the writer during the recording sessions for each of the four shapes.

Task. Participants were seated in front of a computer screen in a quiet room. They listened to the sounds through Sennheiser HD-650 headphones. The graphical interface was designed with the software Max/MSP. The experiment comprised a session of eight trials, that is, four trials with the recorded sounds and four trials with the synthesized sounds. Trials were randomized across participants. In each trial, four shapes and four sound icons were displayed on the computer screen. Subjects were asked to associate each shape to the sound they believed was produced when the shape was drawn. Each sound could be associated to one shape only and vice versa. In practice, the four shapes were displayed on the right-hand side of the screen and the four icons representing the sounds on the left hand side. The shapes were always displayed in the same order. The sounds, which were represented by identical icons, were randomly permuted for each trial. Subjects carried out the association task by moving the sound icon next to the corresponding shape with the computer mouse. They were informed that the sounds had been recorded from a writer drawing each of the four shapes fluently, without lifting the pen. No time limitation was imposed and the subjects could listen to the sounds as often as they wished. For each trial, four sound/shape associations were collected.

Data Analysis

The associations were rated as 1 if the sound was associated to the correct shape and as 0 otherwise. Then, for each subject and each type of sound (synthesized and recorded), the values were averaged across trials to compute an association matrix containing the scores of the association task. In the following, we define the score of success, that is, the score of a correct sound/shape association, as the diagonal values of an association matrix.

To examine the subjects' errors, we compared the scores of the associations between a given sound and the four shapes, and between a given shape and the four sounds. Paired two-tailed *t*-tests were performed on the score of success and the six other scores of association. In addition, for each score, a test of conformity to a standard was carried out with threshold corresponding to 25% of success (equal to the chance threshold).

To evaluate if the task was executed similarly for recorded and synthesized sounds, we carried out a global comparison of the association matrices of the two sound types. To do this, we computed the cophenetic distances between shapes for each matrix and we performed a Pearson's correlation test on these distances. In practice, the analysis was carried out as follows: For each type of sound, we determined, first, a so-called "dissociation" matrix, D^*, defined by: $D^* = 1 - S^*$, where S^* is the symmetrized version of the average association matrix S^* A pairwise "distance" matrix, D, was determined from D^* by choosing the euclidean metric. A hierarchical clustering analysis of D (complete linkage) was then carried out. The cophenetic distances were computed from the resulting dendrogram and were assembled into a vector. The cophenetic distances corresponded to the distances between the shapes estimated at all nodes of the dendrogram.[2] Then, to compare the matrices, the two vectors of cophenetic distances were submitted to a Pearson's correlation test.

Finally, a more precise comparison of the scores of recorded and synthesized sounds was carried out by performing two-sided Wilcoxon signed-ranks test on the rates of success for each shape with the type of sound (recorded vs. synthesized) as a factor.

For all statistical analyses, effects were considered significant if the *p* value was equal to or less than .05. All *p* values were adjusted (Bonferroni correction) for multiple testing.

Results

Table 2 presents the association matrix averaged across subjects for each type of sound. For all sounds, the score of success was significantly above chance ($p < .001$ for each shape, $d_{circle} = 13.19$; $d_{ellipse} = 2.01$; $d_{arches} = 1.91$; $d_{line} = 9.42$ for recorded sounds, $d_{circle} = 13.19$; $d_{ellipse} = 2.81$; $d_{arches} = 2.23$; $d_{line} = 4.9$ for synthesized sounds) and was higher than 80% (highest scores for the line and the circle with almost 100%). Moreover, the scores of success differed significantly from the three other association scores ($p < .001$ for all comparisons). These results revealed that the four sounds had been associated correctly to the corresponding shapes.

Results also showed that, based on the cophenetic distances, the matrices for recorded and synthesized sounds were strongly correlated, $r(4) = 0.89$; $p < .05$. Moreover, the Wilcoxon's tests showed that success rates did not differ between recorded and synthesized sounds for each shape (Circle: $z = 0$; $p = 1$; Ellipse: $z = -1.265$; $p = .21$; Arches: $z = -.632$; $p = .52$; Line: $z = -1.13$; $p = .25$). This revealed that the two types of sound provided similar association scores.

Discussion

The results of this experiment showed that the subjects were able to associate a given friction sound (selected among four) to the correct shape. The scores of success were high for all shapes. Furthermore, the scores obtained for synthesized and recorded sounds did not differ significantly, although the synthesized sounds were modulated by the velocity profiles only. In fact, the two types of sounds differed by the fact that other variables, such as the pressure, the orientation of the mine of the pen and the irregularity of the roughness of the rubbed surface, were implicitly contained in recorded sounds. However, despite of these additional features, the scores were not higher for recorded sounds. This consideration is in line with the study by Schomaker and Plamondon (1990), revealing that no general biological relation exists between these additional features and the kinematic characteristics of a drawn shape. In any case, as the subjects did not draw the shapes themselves, they could not have established any relation between potential acoustical cues linked to pen pressure or angle and the geometry to improve their discrimination of the stimuli.

The result obtained supports the assumption that the velocity profile is perceptually relevant and seems to convey pertinent

[2] This type of analysis is used in phylogenetics to evaluate similarities between matrices and to make a comparison between, for instance, an "empirical" classification computed from macroscopic observations and an "objective" one computed from DNA sequences (Sokal & Rohlf, 1962).

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Table 2
Association Matrices for the Experiment 2

Shape	Circle	Ellipse	Arches	Line
		Recorded sounds		
Circle	**98.75*****	1.25	—	—
	1.25	1.25	—	—
Ellipse	1.25	**81.25*****	17.5	—
	1.25	6.25	6.31	—
Arches	—	17.5	**80*****	2.5
	—	6.31	6.44	1.72
Line	—	—	2.5	**97.5*****
	—	—	3.51	1.72
		Synthesized sounds		
Circle	**98.75*****	1.25	—	—
	1.25	1.25	—	—
Ellipse	—	**87.5*****	12.5	—
	—	4.97	4.97	—
Arches	—	11.25	**82.50*****	6.25
	—	4.62	5.76	3.08
Line	1.25	—	5	**93.75*****
	1.25	—	2.92	3.08

Note. Mean scores and SE in percentage for each shape with recorded (top) and synthesized sounds (bottom). Null values are noted by "—." Significance of the comparison to chance test: $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

information for shape identification. In particular, distinct events such as the silences (corresponding to the zero-crossings in the velocity profile) might be one of the causes of the high recognition scores. From a cognitive point of view, we assume that the association between sounds and visual shapes is enabled based on an internalized model of the gesture evoked by the visual depiction and the perception of sound variations according to the velocity profile of the gesture.

As a matter of fact, in this experiment, the high scores of success were obtained on a set of shapes that were quite distinct as regards the presence (or absence) of cusps and consequently, as regards the underlying movement involved in the drawing of them. In the following experiment, our aim was to examine whether the sound/shape association could also be successfully achieved with shapes that are similar to each other, geometrically, and with acoustic cues based mainly on continuous variations in sound timbre, hence we excluded acoustical cues like silences. More similar shapes should imply more similar velocity profiles that may, in turn, produce friction sounds that might be more difficult to differentiate.

Experiment 3

Method

The procedure and data analysis were the same as in Experiment 2.

Subjects. Eighteen participants took part in the experiment, 8 women and 10 men. Their average age was 31.56 years (SD = 13.73). None of the subjects were familiar with the topic of the study prior to the test. Seven of these subjects participated in Experiment 1, and 17 subjects in Experiment 2.

Stimuli

Static shapes. As for Experiment 2, preliminary informal tests were carried out to choose the shape corpus and the associated sounds based on geometrical and perceptual criteria. In particular, we considered a set of shapes without cusps on the supposition that the corresponding friction sounds would be less distinguishable from a perceptual point of view. Therefore, we kept the circle and the ellipse shapes from the corpus of Experiment 2 and replaced the shapes presenting cusps, that is, arches and lines by loops and lemniscates that do not contain cusps. The four selected shapes are presented in Figure 3.

Recorded sounds. The stimuli corresponding to the circles and ellipses were conserved from Experiment 2. For the loops and lemniscates, the recording sessions took place in the same conditions and with the same writer as in Experiment 2. Sequences corresponding to four periods of sound recordings were selected on the basis of geometrical and temporal characteristics of the writer's performances, described in Table 3. Figure 4 presents the velocity profiles of the four shapes. It is noticeable that these profiles are more similar to each other than those in Experiment 2, which should imply more perceptual confusion between the associated sounds. Only the circle seems to have a velocity profile easily distinguishable from the others.

Synthesized sounds. Friction sounds were synthesized using the same friction model as in Experiments 1 and 2, and by using the velocity profiles collected from the writer during the recording sessions.

Results

The results, presented in Table 4, revealed that, for both recorded and synthesized sounds, all shapes were associated to the correct sound with scores higher than statistical chance; loops with recorded sounds, however, proved an exception to this with a success rate of 29.17% ($p < .001$ for each shape excepted the recorded loops, $d_{circle} = 6.12$; $d_{ellipse} = 0.53$; $d_{lemniscate} = 1.26$; $d_{loops} = 0.13$ for recorded sounds, $d_{circle} > 100$; $d_{ellipse} = 1.30$;

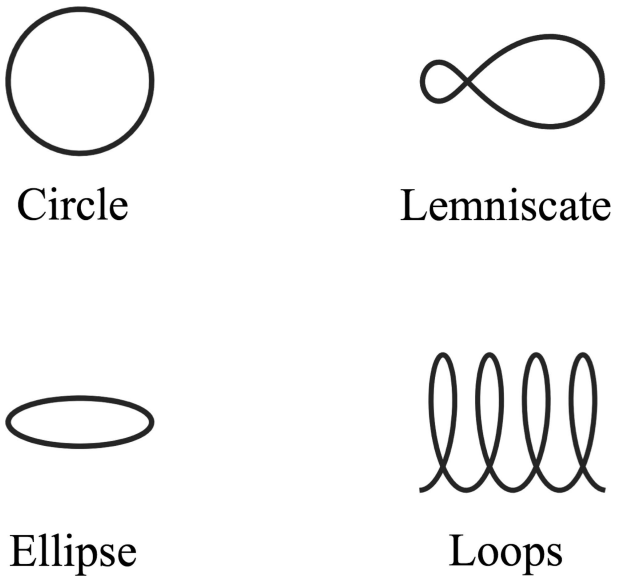


Figure 3. The four shapes used in Experiment 3. For loops, four periods are presented.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Table 3
Geometrical and Temporal Characteristics of the Performances Chosen for the Stimuli of Experiment 3

Shape	Length (cm)	Duration (s)
Circle	62.5	5.2
Ellipse	89.32	5.8
Lemniscate	145.46	5.6
Loops	92.1	5.4

Note. The lemniscate appears to be longer than the three other shapes, but the duration of the associated sound is almost the same as the three other ones because the mean velocity is higher for this shape. The recorded and synthesized stimuli of Experiment 2 are available in supplementary material online.

$d_{lemniscate} = 2.23$; $d_{loops} = 0.7$ for synthesized sounds). The circle and the lemniscate presented the highest scores of success and no confusion with other shapes was observed.

In contrast, results revealed some confusion between ellipses and loops; their scores of association were above statistical chance and did not differ significantly from each other in both directions of association: (a) loop sound with ellipse shape, and (b) ellipse sound with loop shape. Confusion was observed for both recorded (51.39% and 45.83%) and synthesized (45.83% and 43.06%) sounds. These association scores did not differ significantly from the score of success for the ellipses and loops for both recorded sounds, $t(17) = 0.73$; $p = .47$ and $t(17) = 1.89$; $p = .22$, and synthesized sounds, $t(17) = 0.77$; $p = .66$ and $t(17) = 0$; $p = 1$.

Results also revealed that the association matrices for recorded and synthesized sounds were significantly correlated, $r(4) = .94$, $p < .001$. Moreover, the Wilcoxon's tests showed that the scores of success did not differ between the recorded and synthesized sounds of each shape (Circle: $z = -1$; $p = 1$; Ellipse: $z = -.99$; $p = .32$; Lemniscate: $z = -1.348$; $p = .18$; Loops: $z = -1.29$; $p = .20$). This revealed, therefore, that the two types of sounds provided similar association scores.

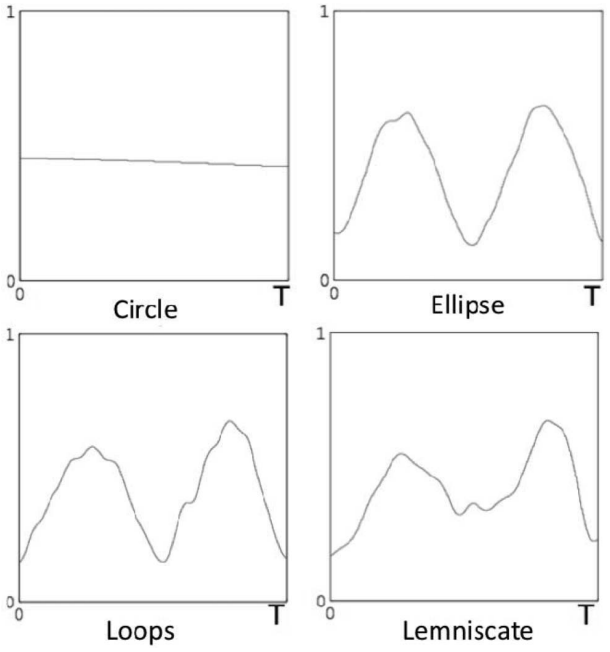


Figure 4. One period of the velocity profiles of the four shapes used in Experiment 3. The velocity profiles are low pass filtered at 10 Hz and normalized in amplitude.

Table 4
Association Matrices of the Experiment 3

Shapes	Circle	Ellipse	Lemniscate	Loops
		Recorded sounds		
Circle	**97.22*****	2.78	—	—
	2.78	2.78	—	—
Ellipse	—	**41.67*****	6.94	**51.39*****
	—	7.29	2.72	6.22
Lemniscate	2.78	9.72	**68.06*****	19.44
	2.78	3.58	8.04	6.87
Loops	—	**45.83*****	25	29.17
	—	7.89	7.83	7.36
		Synthesized sounds		
Circle	**100*****	—	—	—
	0	—	—	—
Ellipse	—	**50*****	4.17	**45.83*****
	—	4.52	2.26	5.05
Lemniscate	—	6.94	**81.94*****	11.11
	—	2.72	6.00	7.46
Loops	—	**43.06*****	13.89	**43.06*****
	—	4.87	5.04	6.00

Note. Mean scores and SE in percentage for each shape with recorded (top) and synthesized sounds (bottom). Null values are noted by "—." Significance of the comparison to chance test: $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

Discussion

The results of this experiment confirmed some of those obtained in Experiment 2: the circle was still perfectly associated with the correct sound (for both recorded and synthesized sounds) even given the three other sounds that were less distinguishable from a perceptual point of view. However, the scores of success for the ellipse were lower than those obtained in Experiment 2. Furthermore, these results revealed some confusion between the ellipse and loops in the shape/sound and sound/shape associations. The lemniscate was well associated with high scores but was marginally confounded with loops and ellipse (scores of association were lower than chance threshold). These confusions can be explained by the proximity of their velocity profiles (see Figure 4). Finally, there was no difference between the scores and confusions elicited by the recorded sounds and those elicited by the synthesized ones. Although the score differences between the recorded and synthesized sounds were not significant, the synthesized sounds always yielded higher scores than the recorded ones, especially for the ellipse, the lemniscate and the loops, for which the score difference was greater than 10%.

These results reinforce the assumption, made in Experiment 2, that the velocity profile constitutes appropriate perceptual information for the association task and the processing of the underlying gesture. However, some confusion was observed in Experiment 3 between the ellipse and loops, for both recorded and synthesized sounds. This indicates that these sounds do not contain enough information on the

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

drawn shapes to distinguish those with similar geometries. By highlighting this confusion, this experiment revealed some limitations to the possibility of a direct relation between sound and shape. This result is also in accordance with the 1/3 power law: to imagine the underlying gesture from the friction sound, we extract kinematics information (velocity profile) and we associate accelerations and decelerations to the curvature characteristics of the drawn shape. Hence, if shapes have similar geometries, the velocity profiles are also similar and the associated sounds are less distinguishable from the auditory point of view.

General Discussion

In the series of experiments presented here, we investigated the human ability to perceive biological movements through sounds and furthermore, to retrieve the drawn shapes from the sounds in an association task. To our knowledge, this is the first time that this topic has been formally addressed. The Experiment 1 investigated whether the velocity profile is a relevant feature for auditory perception using a protocol close to one of Viviani and Stucchi (1992) in the visual modality. To this end, we used a real-time synthesis model to generate friction sounds in which timbre variations were modulated by the velocity profile. This enables to investigate the perceptual relevance of this parameter only. We concluded that these timbre variations should be generated by a velocity profile that obeys the 1/3 power law to evoke a natural and fluid biological movement. This study revealed that sounds can adequately inform about human movements if their dynamical acoustic characteristics are in accordance with the way the movements are performed.

The Experiments 2 and 3 further investigated whether participants were able to extract the spatial characteristics of visual shapes from the sounds and we opted for an association task. In Experiment 2, we compiled a sound corpus assumed to be perceptually easy to discriminate. In Experiment 3, the corpus comprised less discriminable stimuli. As expected, high scores of success for distinct shapes and some confusion for similar shapes were observed. Discriminating between visual shapes on the basis of their produced sounds is, therefore, possible if the acoustic characteristics of the sounds differ sufficiently. The lack of score differences between recorded and synthesized sounds confirmed that these characteristics are related mainly to the velocity profile of the underlying drawing movement, which complies with the 1/3 power law.

From an ecological perspective, the specific pattern of timbre variations induced by the velocity profile can be considered as a transformational invariant enabling the evocation of the underlying drawing movement. Such timbre variations afford the action of drawing fluidly and naturally, and this result is in line with many studies dealing with auditory perception of an acoustical event (Gaver, 1993b; Li, Logan and Pastore, 1991; Repp, 1987; Warren & Verbrugge, 1984). To complement the ecological framework we will also discuss the results according to the ideomotor framework (Hommel et al., 2001; Prinz, 1997). In a recent study by Young et al. (2013), subjects were asked to reproduce walking patterns from walking sounds in real time with different stride lengths, and to discriminate these stimuli in perceptual tasks. They observed that the characteristics of the reproduced walking patterns were similar to the target ones. Moreover, the subjects were able to perform the tasks accurately even when only the kinematic information was present in the synthesized target walking sounds. This suggests that the auditory perception of an action seems to activate the same motor schemes as when we act, and finally that, listening to an action enables one to imagine it, and even, to reproduce it in real time. The authors proposed that common cognitive representations are involved both in the perception and in the planning of an action, as suggested in recent studies which argued in favor of a reunification of direct and indirect perception (Cisek & Kalaska, 2010; Norman, 2002). In the case of their experiments, the authors assumed that the tasks were well accomplished thanks to an audio-motor unified percept. We propose the same conceptual processing of sensory information in the perception of human drawing through sounds. In particular, inferring drawn shapes from sounds requires an internalized association between these two percepts. This relation is not obvious because graphical shapes are not commonly associated to sounds. However, based on results from Experiment 1 and from studies found in the literature (cf. Introduction), we suggest that both a visual shape and a sound can evoke a given human gesture and that the relevant information is contained in the velocity profile as conveyed by the movement of a visual dot or by timbre variations of the sound. We propose that the evoked human gesture serves as a medium to accomplish the task of associating a shape and a sound thanks to an amodal representation of biological gesture, as proposed in Viviani, Baud-Bovy, and Redolfi (1997). Indeed, we were unable to conclude whether subjects deduced the shape from the sound or the sound from the shape. The sensorimotor representation of biological movement constitutes the central point of such a process. This can be reinforced by the results from Viviani and Stucchi (1989), which showed that the link between the dynamical representation of the shape (a moving spot) and the drawn (static) shape is mediated in a similar way by a sensorimotor representation of the underlying movement.

Hence, we speculate that sensory information (visual, auditory, and sensorimotor) can be integrated to provide a unified percept of the "drawing event." This conception also suggests that common motor rules constrain the perception of both auditory and visual biological movements, thus confirming previous results on the existence of motor-perceptual relations. As suggested by Young et al. (2013), these results are also in line with motor theories of speech perception, especially with the version assuming a role of motor knowledge in perceptual processes (Viviani & Stucchi, 1992 for a review; Jeannerod, 1995; Zatorre, Chen, & Penhune, 2007; Bangert et al., 2006 in the context of music perception and production). More generally, such sensory integration may rely on a multimodal representation, such as the model proposed by Griffiths and Warren (2004). Indeed, substantial literature pertaining to such cognitive models does not exist at this time.

Conclusion and Perspectives

This study demonstrated that the friction sounds produced when someone is drawing are sufficiently informative to evoke the underlying gesture, and to a certain extent the drawn shape. We focused on the kinematic information (i.e., the velocity profile) using a synthesis process and showed that it is possible to calibrate the timbre of a sound such that the evoked motion corresponds to a biological movement with a velocity profile that matches the 1/3 power law. At last, two experiments pointed out that shapes can be retrieved from friction sounds, and that this discrimination was enabled when only the kinematic information was present in the sound. The main results of

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

this study are twofold: first, from an ecological perspective, the velocity profile can be considered as a transformational invariant as it affords the recognition of human movements from the generated friction sounds. Second, a relation is enabled between drawn shapes and produced sounds, which strongly suggests that common rules constrain perception and action of biological motions.

Several perspectives can be highlighted at this point. First, it would be of interest to investigate more globally and accurately, the relations between sound, visual shape, and evoked movement. In particular, the auditory neural correlates of the sensorimotor representation of biological movements have already been investigated and it has been shown that those related to the 1/3 power law of motion are much stronger than other types of motion (Dayan et al., 2007). As revealed by the experiments reported here, synthesis is an efficient way to investigate the brain correlates of the auditory modality because any velocity profile (respecting or not the 1/3 power law) can be easily generated by this procedure. Moreover, Lewis et al. (2004) showed that certain cortical areas (posterior portions of the middle temporal gyri) are involved in both visual biological motion and sound perception and might participate in the audio-visual integration process.

Second, in Experiments 2 and 3, participants associated a given sound to a given shape among a limited number of visually displayed shapes. To access the representation of the drawn visual shape directly from the sound, it would be interesting to conduct an identification task in which participants are asked to draw the shape evoked by a sound without any visual reference and to compare both the drawn shape and the kinematics of the drawing movement with the real ones. A parameterized graphical user interface based on an interactive synthesis tool could be used for this purpose (cf. Merer et al., 2013). However, the confusions observed between similar shapes in Experiment 3 support that retrieving the correct drawn shape without any visual model may prove somewhat difficult. Another interesting perspective would be to carry out the same experiments, but with the recordings of each subject, to investigate whether idiosyncratic knowledge of pressure and pen angle behaviors could have improved shape recognition for those shapes that were confounded. It might, therefore, be of interest to exaggerate the values of the synthesis parameters, particularly those related to the velocity profile, to try to boost shape identification. This possibility may be useful in many applications for sound design, sonification, or even for musical purposes.

References

Aramaki, M., Besson, M., Kronland-Martinet, R., & Ystad, S. (2011). Controlling the perceived material in an impact sound synthesizer. *Audio, Speech, and Language Processing, 19,* 301–314. doi:10.1109/TASL.2010.2047755

Bangert, M., Peschel, T., Schlaug, G., Rotte, M., Drescher, D., Hinrichs, H., . . . Altenmüller, E. (2006). Shared network for auditory and motor processing in professional pianists: Evidence from fMRI conjunction. *Neuroimage, 30,* 917–926. doi:10.1016/j.neuroimage.2005.10.044

Carello, C., Anderson, K. L., & Kunkler-Peck, A. J. (1998). Perception of object length by sound. *Psychological Science, 9,* 211–214. doi:10.1111/1467-9280.00040

Carlyon, R. P., Macherey, O., Frijns, J. H. M., Axon, P. R., Kalkman, R. K., Boyle, P., . . . Dauman, R. (2010). Pitch comparisons between electrical stimulation of a cochlear implant and acoustic stimuli presented to a normal-hearing contralateral ear. *Journal of the Association for Research in Otolaryngology, 11,* 625–640. doi:10.1007/s10162-010-0222-7

Castiello, U., Giordano, B. L., Begliomini, C., Ansuini, C., & Grassi, M. (2010). When ears drive hands: The influence of contact sound on reaching to grasp. *PLoS One, 5,* e12240. doi:10.1371/journal.pone.0012240

Cisek, P., & Kalaska, J. F. (2010). Neural mechanisms for interacting with a world full of action choices. *Annual Review of Neuroscience, 33,* 269–298. doi:10.1146/annurev.neuro.051508.135409

Dayan, E., Casile, A., Levit-Binnun, N., Giese, M. A., Hendler, T., & Flash, T. (2007). Neural representations of kinematics laws of motion: Evidence for action-perception coupling. *Proceedings of the National Academy of Sciences of the United States of American, 104,* 20582–20587. doi:10.1073/pnas.0710033104

Frémiot, M., Mandelbrojt, J., Formosa, M., Delalande, G., Pedler, E., Malbosc, P., & Gobin, P. (1996). Les Unités Sémiotiques Temporelles: éléments nouveaux d'analyse musicale. *diffusion ESKA*. MIM Laboratoire Musique et Informatique de Marseille, documents musurgia edition [The temporal semiotic unities: new element of musical analyses].

Freyd, J. J. (1983a). Representing the dynamics of a static form. *Memory & Cognition, 11,* 342–346. doi:10.3758/BF03202447

Freyd, J. J. (1983b). The mental representation of movement when static stimuli are viewed. *Perception & Psychophysics, 33,* 575–581. doi:10.3758/BF03202940

Gaver, W. W. (1993a). What in the world do we hear?: An ecological approach to auditory event perception. *Ecological Psychology, 5*(1), 1–29. doi:10.1207/s15326969eco0501_1

Gaver, W. W. (1993b). How do we hear the world? Explanations in ecological acoustics. *Ecological Psychology, 5,* 285–313. doi:10.1207/s15326969eco0504_2

Gibson, J. J. (1966). *The senses considered as perceptual systems,* Boston, MA: Houghton Mifflin

Giordano, B. L., & McAdams, S. (2006). Material identification of real impact sounds: Effects of size variation in steel, glass, wood and plexiglass plates. *Journal of the Acoustical Society of America, 119,* 1171–1181. doi:10.1121/1.2149839

Giordano, B. L., Rocchesso, D., & McAdams, S. (2010). Integration of acoustical information in the perception of impacted sound sources: The role of information accuracy and exploitability. *Journal Experimental Psychology: Human perception and Performance, 36,* 462–476. doi:10.1037/a0018388

Grassi, M. (2005). Do we hear size or sound: Balls dropped on plates. *Perception & Psychophysics, 67,* 274–284. doi:10.3758/BF03206491

Grassi, M., Pastore, M., & Lemaitre, G. (2013). Looking at the world with your ears: How do we get the size of an object from its sound? *Acta Psychologica, 143,* 96–104. doi:10.1016/j.actpsy.2013.02.005

Griffiths, T. D., & Warren, J. D. (2004). What is an auditory object? *Nature Reviews Neuroscience, 5,* 887–892. doi:10.1038/nrn1538

Hommel, B. (2004). Event files: Feature binding in and across perception and action. *Trends in Cognitive Sciences, 8,* 494–500. doi:10.1016/j.tics.2004.08.007

Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The theory of event coding (TEC): A framework for perception and action planning, *Behavioral and Brain Sciences, 24,* 849–878. doi:10.1017/S0140525X01000103

Honing, H. (2003). The final ritard: On music, motion, and kinematic models. *Computer Music Journal, 27,* 66–72. doi:10.1162/014892603322482538

Houben, M. M. J., Kohlrausch, A., & Hermes, D. J. (2004). Perception of the size and speed of rolling balls by sound. *Speech Communication, 43,* 331–345. doi:10.1016/j.specom.2004.03.004

Houben, M. M. J., Kohlrausch, A., & Hermes, D. J. (2005). The contribution of spectral and temporal information to the auditory perception of the size and speed of rolling balls. *Acta acustica united with Acustica, 91,* 1007–1015.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

James, K. H., & Gauthier, I. (2006). Letter processing automatically recruits a sensory? Motor brain network. *Neuropsychologia, 44,* 2937–2949. doi:10.1016/j.neuropsychologia.2006.06.026

Jeannerod, M. (1995). Mental imagery in the motor context. *Neuropsychologia, 33,* 1419–1432. doi:10.1016/0028-3932(95)00073-C

Johnson, M. L., & Larson, S. (2003). Something in the way she moves, metaphors of musical motion, *Metaphor and Symbol, 18,* 63–84. doi:10.1207/S15327868MS1802_1

Klatzky, R. L., Pai, D. K., & Krotkov, E. P. (2000). Perception of material from contact sounds. *Presence: Teleoperators & Virtual Environments, 9,* 399–410. doi:10.1162/105474600566907

Kohler, E., Keysers, C., Umilta, M. A., Fogassi, L., Gallese, V., & Rizzolatti, G. (2002). Hearing sounds, understanding actions: Action representation in mirror neurons. *Science, 297,* 846–848. doi:10.1126/science.1070311

Kunkler-Peck, A. J., & Turvey, M. T. (2000). Hearing shape. *Journal of Experimental Psychology: Human Perception and Performance, 26,* 279–294. doi:10.1037/0096-1523.26.1.279

Lacquaniti, F., Terzuolo, C., & Viviani, P. (1983). The law relating the kinematic and figural aspects of drawing movements. *Acta Psychologica, 54,* 115–130. doi:10.1016/0001-6918(83)90027-6

Lakatos, S., McAdams, S., & Caussé, R. (1997). The representation of auditory source characteristics: Simple geometric form. *Perception & Psychophysics, 59,* 1180–1190. doi:10.3758/BF03214206

Lemaitre, G., & Heller, L. M. (2012). Auditory perception of material is fragile, while action is strikingly robust. *Journal of the Acoustical Society of America, 131,* 1337–1348. doi:10.1121/1.3675946

Lewis, J. W., Wightman, F. L., Brefczynski, J. A., Phinney, R. E., Binder, J. R., & DeYoe, E. A. (2004). Human brain regions involved in recognizing environmental sounds. *Cerebral Cortex, 14,* 1008–1021. doi:10.1093/cercor/bhh061

Li, X., Logan, R. J., & Pastore, R. E. (1991). Perception of acoustic source characteristics: Walking sounds. *Journal of the Acoustical Society of America, 90,* 3036–3049. doi:10.1121/1.401778

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74,* 431–461. doi:10.1037/h0020279

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition, 21,* 1–36. doi:10.1016/0010-0277(85)90021-6

Longcamp, M., Anton, J. L., Roth, M., & Velay, J. L. (2003). Visual presentation of single letters activates a premotor area involved in writing. *Neuroimage, 19,* 1492–1500. doi:10.1016/S1053-8119(03)00088-0

Longcamp, M., Boucard, C., Gilhodes, J. C., Anton, J. L., Roth, M., Nazarian, B., & Velay, J. L. (2008). Learning through hand-or typewriting influences visual recognition of new graphic shapes: Behavioral and functional imaging evidence. *Journal of Cognitive Neuroscience, 20,* 802–815. doi:10.1162/jocn.2008.20504

Longcamp, M., Tanskanen, T., & Hari, R. (2006). The imprint of action: Motor cortex involvement in visual perception of handwritten letters. *Neuroimage, 33,* 681–688. doi:10.1016/j.neuroimage.2006.06.042

McAdams, S. (1993). Recognition of sound sources and events. In S. McAdams & E. Bigand (Eds.), *Thinking in sound: The cognitive psychology of human audition* (pp. 146–198). New York, NY: Oxford University Press.

McAdams, S., Chaigne, A., & Roussarie, V. (2004). The psychomechanics of simulated sound sources: Material properties of impacted bars. *Journal of the Acoustical Society of America, 115,* 1306–1320. doi:10.1121/1.1645855

Merer, A., Aramaki, M., Ystad, S., & Kronland-Martinet, R. (2013). Perceptual characterization of motion evoked by sounds for synthesis control purposes. *ACM Transaction on Applied Perception (TAP), 10,* 1. doi:10.1145/2422105.2422106

Merer, A., Ystad, S., Kronland-Martinet, R., & Aramaki, M. (2008). Semiotics of sounds evoking motions: Categorization and acoustic features. In R. Kronland-Martinet, S. Ystad, & K. Jensen (Eds.), *CMMR 2007. Sense of sounds* (pp. 139–158). Berlin, Germany: Springer, LNCS.

Norman, J. (2002). Two visual systems and two theories of perception: An attempt to reconcile the constructivist and ecological approaches. *Behavioral and Brain Sciences, 25,* 73–96.

Prinz, W. (1997). Perception and action planning. *European Journal of Cognitive Psychology, 9,* 129–154. doi:10.1080/713752551

Repp, B. H. (1987). The sound of two hands clapping: An exploratory study. *Journal of the Acoustical Society of America, 81,* 1100–1109. doi:10.1121/1.394630

Richardson, M. J. E., & Flash, T. (2002). Comparing smooth arm movements with the two-thirds power law and the related segmented-control hypothesis. *The Journal of Neuroscience, 22,* 8201–8211.

Schomaker, L. R. B., & Plamondon, R. (1990). The relation between pen force and pen-point kinematics in handwriting. *Biological Cybernetics, 63,* 277–289. doi:10.1007/BF00203451

Sokal, R. R., & Rohlf, F. J. (1962). The comparison of dendrograms by objective methods. *Taxon, 11,* 33–40. doi:10.2307/1217208

Van den Doel, K., Kry, P. G., & Pai, D. K. (2001). FoleyAutomatic: Physically-based sound effects for interactive simulation and animation. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques ACM,* 537–544.

Viviani, P. (2002). Motor competence in the perception of dynamic events: A tutorial. In W. Prinz & B. Hommel (Eds.), *Common mechanisms in perception and action* (pp. 406–442). New York, NY: Oxford University Press.

Viviani, P., Baud-Bovy, G., & Redolfi, M. (1997). Perceiving and tracking kinesthetic stimuli: Further evidence of motor-perceptual interactions. *Journal of Experimental Psychology: Human Perception and Performance, 23,* 1232–1252. doi:10.1037/0096-1523.23.4.1232

Viviani, P., & Flash, T. (1995). Minimum-jerk, two-thirds power law, and isochrony: Converging approaches to movement planning. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 32–53. doi:10.1037/0096-1523.21.1.32

Viviani, P., & McCollum, G. (1983). The relation between linear extent and velocity in drawing movements. *Neuroscience, 10,* 211–218. doi:10.1016/0306-4522(83)90094-5

Viviani, P., & Stucchi, N. (1989). The effect of movement velocity on form perception: Geometric illusions in dynamic displays. *Perception & Psychophysics, 46,* 266–274. doi:10.3758/BF03208089

Viviani, P., & Stucchi, N. (1992). Biological movements look uniform: Evidence of motor-perceptual interactions. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 603–623. doi:10.1037/0096-1523.18.3.603

Viviani, P., & Terzuolo, C. (1982). Trajectory determines movement dynamics. *Neuroscience, 7,* 431–437. doi:10.1016/0306-4522(82)90277-9

Warren, W. H., & Verbrugge, R. R. (1984). Auditory perception of breaking and bouncing events: A case study in ecological acoustics. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 704–712. doi:10.1037/0096-1523.10.5.704

Wildes, R. P., & Richards, W. A. (1988). Recovering material properties from sound. In W. A. Richards (Ed.), *Natural computation* (pp. 356–363). Cambridge, MA: MIT Press.

Young, W., Rodger, M., & Craig, C. M. (2013). Perceiving and reenacting spatiotemporal characteristics of walking sounds. *Journal of Experimental Psychology: Human Perception and Performance, 39,* 464–476. doi:10.1037/a0029402

Zatorre, R. J., Chen, J. L., & Penhune, V. B. (2007). When the brain plays music: Auditory-motor interactions in music perception and production. *Nature Reviews Neuroscience, 8,* 547–558. doi:10.1038/nrn2152

Received January 28, 2013
Revision received November 6, 2013
Accepted November 15, 2013 ■

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Sound Categorization and Conceptual Priming for Nonlinguistic and Linguistic Sounds

Mitsuko Aramaki[1,2], Céline Marie[1,2], Richard Kronland-Martinet[3], Sølvi Ystad[3], and Mireille Besson[1,2]

Abstract

■ The aim of these experiments was to compare conceptual priming for linguistic and for a homogeneous class of nonlinguistic sounds, impact sounds, by using both behavioral (percentage errors and RTs) and electrophysiological measures (ERPs). Experiment 1 aimed at studying the neural basis of impact sound categorization by creating typical and ambiguous sounds from different material categories (wood, metal, and glass). Ambiguous sounds were associated with slower RTs and larger N280, smaller P350/P550 components, and larger negative slow wave than typical impact sounds. Thus, ambiguous sounds were more difficult to categorize than typical sounds. A category membership task was used in Experiment 2. Typical sounds were followed by sounds from the same or from a different category or by ambiguous sounds. Words were followed by words, pseudowords, or nonwords. Error rate was highest for ambiguous sounds and for pseudowords and both elicited larger N400-like components than same typical sounds and words. Moreover, both different typical sounds and nonwords elicited P300 components. These results are discussed in terms of similar conceptual priming effects for nonlinguistic and linguistic stimuli. ■

INTRODUCTION

Word processing is specific in that the sequences of phonemes that form spoken words (or the sequence of graphemes that form written words) are not meaningful by themselves but acquire meaning through the process of double articulation. Thus, there is generally no direct relationship between the sound (or form) of a word and its meaning (de Saussure, 1916). By contrast, a causal relationship exists between the perceptual characteristics of environmental sounds (e.g., broken glass) and the meaning that is derived from them (e.g., a glass was broken; Ballas, 1993; Ballas & Howard, 1987). On the basis of such differences, one may expect words and environmental sounds to be processed differently. However, results of several experiments that have used the ERP method to address this question argue in favor of the similarities rather than the differences in the processing of words and environmental sounds (Orgs, Lange, Dombrowski, & Heil, 2006, 2007, 2008; Cummings et al., 2006; Plante, Van Petten, & Senkfor, 2000; Van Petten & Rheinfelder, 1995).

One of the first studies of conceptual priming was conducted by Van Petten and Rheinfelder (1995). In their first experiment, environmental sounds were used as primes and related words, unrelated words, and pseudowords were used as targets. Participants were asked to decide whether the target stimulus was a word or not (lexical decision task). Results showed higher error rate for pseudowords than for words and faster RTs for related than for unrelated words. Thus, these results demonstrated conceptual priming between environmental sounds and words. To examine the neural basis of this effect, they took advantage of the N400 component (Kutas & Hillyard, 1980) to compare priming when the prime is an environmental sound and the target a related or an unrelated printed word and vice versa. Results revealed that conceptual priming, as reflected by the N400 effect (i.e., the difference between unrelated and related targets), was very similar for environmental sounds and for words. This result argues in favor of the similarity of the neural processes involved in computing the meaning of words and environmental sounds. However, although the typical, slightly larger over the right than left hemisphere distribution of the N400 effect was found for target words (Kutas, Van Petten, & Besson, 1988; Kutas & Hillyard, 1982), the N400 effect to environmental sounds was larger over the left hemisphere. Similar interhemispheric differences in the N400 effect were reported in a subsequent study (Plante et al., 2000) in which priming for pairs of line drawing and environmental sounds, on one side, and for pairs of printed and spoken words, on the other side, was compared using the same task as in Van Petten and Rheinfelder (1995).

More recently, Orgs et al. (2006) used printed words as primes followed by environmental sounds as targets (or vice versa). Primes and targets were semantically related or

[1]CNRS—Institut de Neurosciences Cognitives de la Méditerranée, Marseille Cedex, France, [2]Université Aix-Marseille, Marseille Cedex, France, [3]CNRS—Laboratoire de Mécanique et d'Acoustique, Marseille Cedex, France

© 2009 Massachusetts Institute of Technology

unrelated, and participants were asked to decide whether the words and environmental sounds fitted together or not. Results showed slower RTs and larger N400 amplitude for unrelated than for related targets. Moreover, the N400 effect showed an earlier onset for environmental sounds than for words over parieto-occipital sites. In subsequent studies with similar design and stimuli, participants were asked to perform either a physical task or a semantic task (Orgs et al., 2007, 2008). In both tasks, the authors found an N400 effect that was taken to reflect the automatic processes that mediate analysis of sound meaning. Finally, Cummings et al. (2006) also used a cross-modal priming design with pictures presented as primes and related or unrelated spoken words, environmental sounds, or nonmeaningful sounds (defined by the authors as "computer-generated sounds that were not easily associated with any concrete semantic concept"; Cummings et al., 2006, p. 104) presented as targets. Participants were asked to decide whether the two stimuli matched or mismatched. They reported an N400 effect for both words and environmental sounds but not for nonmeaningful sounds. Again, the N400 effect had an earlier onset for environmental sounds than for words, but in contrast with Orgs et al. (2006), the N400 effect was larger for environmental sounds than for words over frontal sites (F3/F4). Moreover, in contrast to Plante et al. (2000) and Van Petten and Rheinfelder (1995), they reported no interhemisphere differences. Findings from an experiment using event-related desynchronization to compare words and environmental sounds suggest the involvement of left-lateralized phonological and semantic processes for words and of distributed processes in both hemispheres for environmental sounds (Lebrun et al., 2001). On the basis of these results, Lebrun et al. (2001) suggested a common semantic system for both words and environmental sounds but with more specific perceptual processing for the later.

Other studies have also examined conceptual priming between music and language. Results have shown the occurrence of an N400 component to unrelated visual word targets when primes were long musical excerpts (several seconds; Koelsch et al., 2004). Recently, Daltrozzo and Schön (2009) used either words or short musical excerpts (1 sec) as primes and targets and found an N400 effect in both cases. However, the scalp distribution of the N400 effect was temporal for musical excerpts and parietal for visual word targets. To our knowledge, only one study has examined priming effects for pairs of musical excerpts presented both as prime and as target stimuli (Frey et al., 2009). Although the musical excerpts used as primes conveyed specific concepts, the musical excerpts used as targets either conveyed the same concept as the prime (congruous) or started with the same concept but shifted midstream into another concept (incongruous). Results showed an N400-like component, which was largest over right frontal regions, to incongruous musical excerpts in nonmusicians.

Taken together, results of these studies, which used different experimental designs, stimuli, and tasks (i.e., lexical decision task, matching tasks, physical or semantic priming), nevertheless concur in showing that environmental sounds or musical excerpt targets that are unrelated to word or picture primes elicit increased negative components in the N400 latency band compared with related targets. However, the scalp distribution of these effects varies either between hemispheres (Lebrun et al., 2001; Plante et al., 2000; Van Petten & Rheinfelder, 1995) or along the anterior-posterior dimension (Daltrozzo & Schön, 2009; Cummings et al., 2006; Orgs et al., 2006). Thus, whether conceptual priming for environmental sounds, music, and words rely on similar or different processes still remains an open issue.

The Present Studies

In the different studies summarized above (except for Frey et al., 2009), some form of cross-modal priming was always used between pictures, printed or spoken words, musical excerpts, on one side, and nonlinguistic sounds (environmental sounds, musical excerpts), on the other side. As a consequence, the presence of words or pictures in the experimental design may have encouraged the use of linguistic encoding strategies. Thus, the possibility remains that, within such experimental contexts, participants associated a verbal label to each sound thereby explaining the similarity of the N400 for words and environmental sounds. In a recent study, Schön, Ystad, Kronland-Martinet, and Besson (2009) also used a cross-modal priming design but they tried to minimize linguistic mediation by presenting sounds with no easily identifiable sound sources (i.e., a verbal label could not easily be associated to the sounds). Again, larger N400 components were found for targets (sounds or words) that were unrelated to the primes (words or sounds). However, this cross-modal design still involved words as prime or target stimuli. Therefore, our first aim was to reduce the influence of linguistic mediation by using only nonlinguistic sounds as prime and target. Moreover, in previous studies, the set of nonlinguistic sounds that were used were often very diverse and nonhomogeneous (e.g., animal or human nonspeech sounds, instrumental sounds, and everyday life sounds). As these different types of sounds may engage different processes, our second aim was to use only one homogeneous class of environmental sounds: impact sounds from wood, metal, or glass. Finally, to our knowledge, priming effects including only linguistic primes and linguistic targets on one side and nonlinguistic primes and nonlinguistic targets on the other side have never been directly compared within subjects. Thus, our third aim was to directly compare within-subjects conceptual priming for impact sounds and for linguistic sounds by using the same task with both types of stimuli. However, to use a priming design including only nonlinguistic sounds as stimuli, we first needed to create typical sounds from different impact sound categories (i.e.,wood, metal, glass) and ambiguous sounds. The aims

of Experiment 1 were to create such sounds and to study the neural basis of impact sound categorization.

EXPERIMENT 1: SOUND CATEGORIZATION

To create typical and ambiguous impact sounds, we used a morphing technique. First, sounds from three material categories (wood, metal, and glass) were recorded, analyzed, and resynthesized using an analysis–synthesis method (Aramaki & Kronland-Martinet, 2006) to generate realistic synthetic sounds. Second, sound continua were created that simulate progressive transitions between sounds from different materials (i.e., wood–metal, wood–glass, and metal–glass continua). Although sounds at extreme positions on the continua were synthesized to be as similar as possible to natural sounds, sounds at intermediate positions were synthesized by interpolating the acoustic parameters characterizing sounds at extreme positions. They were consequently ambiguous (e.g., neither wood nor metal). Sounds from the different continua were randomly presented, and participants were asked to categorize each sound as wood, metal, or glass. If sounds at extreme positions of the continua are indeed perceived as typical exemplars of their respective categories, they should be categorized faster and with lower error rate than sounds at intermediate positions on the continua.

Little is known about the neural basis of impact sound perception. To investigate this issue, we also recorded ERPs while participants performed the categorization task. Results of studies on the categorization of nonspeech stimuli have shown that the amplitude of the N200–P300 complex, which typically follows the N100–P200 exogenous complex, is influenced by the difficulty of the categorization task: The N200 component is larger and the P300 component is smaller to stimuli that are more difficult to categorize (Donchin & Coles, 1988; Ritter, Simson, & Vaughan, 1983; Donchin, 1981; Ritter, Simson, Vaughan, & Friedman, 1979; Simson, Vaughan, & Ritter, 1977). Thus, if ambiguous sounds are more difficult to categorize than typical sounds because they are composed of hybrid acoustic features, they should elicit larger N200 and smaller P300 components than typical sounds.

Methods

Subjects

A total of 25 participants were tested in this experiment that lasted for about 1 hour. Three participants were excluded from final data analysis because of the high number of trials contaminated by ocular and muscular artifacts. The remaining 22 participants (11 women and 11 men; 19–35 years old) were all right-handed, were nonmusicians (no formal musical training), had normal audition, and had no known neurological disorders. They all gave written consent to participate to the experiment and were paid for their participation.

Stimuli

We first recorded sounds by impacting everyday life objects made of different materials (i.e., wooden beams, metallic plates, and various glass bowls) to insure the generation of realistic familiar sounds. Then, we used a simplified version of the model described in Aramaki and Kronland-Martinet (2006) on the basis of an additive synthesis technique to resynthesize these recorded sounds (44.1-kHz sampling frequency). From a physical point of view, because the vibrations of an impacted object (under free oscillations) can generally be written as a sum of exponentially damped sinusoids, the recorded sounds are considered to be well described by

$$s(t) = \theta(t) \sum_{m=1}^{M} A_m \sin(2\pi f_m t) e^{-\alpha_m t} \qquad (1)$$

where $\theta(t)$ is the Heaviside unit step function, M is the number of sinusoidal components, and the parameters A_m, f_m, and α_m are the amplitude, frequency, and damping coefficient of the mth component, respectively. The synthesis parameters of this model (i.e., M, A_m, f_m, and α_m) were estimated from the analysis of the recorded sounds (examples of analysis–synthesis processes can be found in Kronland-Martinet, Guillemain, & Ystad, 1997). In practice, many sounds from each material category were recorded and resynthesized and the five most representative sounds per category (as judged by seven listeners) were selected for the current study.

All sounds were tuned to the same chroma (note C that was closest to the original pitch) and were equalized in loudness by gain adjustments. Averaged sound duration was 744 msec for wood, 1667 msec for metal, and 901 msec for glass category. Because the damping is frequency dependent (see Equation 1), the damping coefficient of each tuned component was modified according to a damping law estimated from the original sound (Aramaki, Baillères, Brancheriau, Kronland-Martinet, & Ystad, 2007).

A total of 15 sound continua were created as progressive transitions between two material categories (i.e., 5 different continua for each transition: wood–metal, wood–glass, and glass–metal). Each continuum comprised 20 sounds that were generated using additive synthesis (see Equation 1). Each sound of the continuum was obtained by combining the spectral components of the two extreme sounds and by varying the amplitude and damping coefficients. Amplitude variations were obtained by applying a cross-fade technique between the two extreme sounds. Damping coefficients were estimated by a hybrid damping law resulting from the interpolation between the damping laws of the two extreme sounds. Note that this manipulation allowed creating hybrid sounds that differed from a simple mix between the two extreme sounds because at each step, the spectral components are damped following a same hybrid damping law. All sounds are available at http://www.lma.cnrs-mrs.fr/~kronland/Categorization/sounds.html.

Procedure

A total of 300 sounds were presented in a random order within five blocks of 60 sounds through one loudspeaker (Tannoy S800) located 1 m in front of the participants. They were asked to listen to each sound and to categorize it as wood, metal, or glass as quickly as possible by pressing one button on a three-button response box. The association between response buttons and sounds was balanced across participants. The experiment was conducted in a faradized room. A row of XXXX was presented on the screen 2500 msec after sound onset for 1500 msec to give participants time to blink, and the next sound was then presented after a 500-msec delay.

Recording ERPs

The EEG was recorded continuously from 32 Biosemi Pin-type active electrodes (Amsterdam University) mounted on an elastic headcap and located at standard left and right hemisphere positions over frontal, central, parietal, occipital, and temporal areas (international extended 10/20 system; Jasper, 1958): Fz, Cz, Pz, Oz, Fp1, Fp2, AF3, AF4, F7, F8, F3, F4, Fc5, Fc6, Fc1, Fc2, T7, T8, C3, C4, Cp1, Cp2, Cp5, Cp6, P3, P4, PO3, PO4, P7, P8, O1, and O2. Moreover, to detect horizontal eye movements and blinks, the EOG was recorded from flat-type active electrodes placed 1 cm to the left and right of the external canthi and from an electrode beneath the right eye. Two additional electrodes were placed on the left and right mastoids. EEG was recorded at 512-Hz sampling frequency using Biosemi amplifier. The EEG was rereferenced off-line to the algebraic average of the left and right mastoids and filtered with a band-pass of 0–40 Hz.

Data were analyzed using Brain Vision Analyzer software (Brain Products, Munich), segmented in single trials of 2500 msec starting 200 msec before the onset of the sound and averaged as a function of the type of sound (i.e., typical vs. ambiguous).

Results

Behavioral Data

Participants' responses and RTs were collected for each sound and were averaged across participants. Sounds categorized within a category (wood, metal, or glass) by more than 70% of the participants were considered as typical sounds; sounds categorized within a category by less than 70% of the participants were considered as ambiguous sounds. As can be seen in Figure 1 (top), participants' responses are consistent with the position of the sound on the continua so that typical sounds are located at extreme positions and ambiguous sounds at intermediate positions on the continua.

RTs to typical and ambiguous sounds were submitted to repeated measures ANOVAs (for this and following statistical analyses, effects were considered significant if the p value was equal to or less than .05) that included Type of Sounds (typical vs. ambiguous) and Continua (wood–metal, wood–glass, and glass–metal) as within-subject factors. For typical sounds, only RTs associated to correct responses were taken into account. As shown in Figure 1 (bottom), RTs to typical sounds (984 msec) were shorter than RTs to ambiguous sounds (1165 msec), $F(1, 21) = 74.00, p < .001$. Moreover, the Type of Sounds × Continua interaction was significant, $F(2, 42) = 22.24, p < .001$. Results of post hoc comparisons (Tukey tests) showed that although RTs were shorter for typical than for ambiguous sounds for each continuum ($p < .01$), this difference was larger for the wood–glass continuum (281 msec) than for the wood–metal (184 msec) and glass–metal (79 msec) continua.

Electrophysiological Data

Separate ANOVAs were conducted for midline and lateral electrodes. Type of sounds[1] (typical vs. ambiguous) and Electrodes (Fz, Cz, Pz) were used as factors for midline analyses. Type of Sounds, Hemispheres (left vs. right), ROIs (fronto-central R1, centro-temporal R2, and centro-parietal R3), and Electrodes (three for each ROI: [AF3, F3, FC5]/[AF4, F4, FC6]; [T7, C3, CP5]/[T8, C4, CP6]; and [P7, P3, CP1]/[P8, P4, CP2]) were included for lateral analyses. On the basis of visual inspection of the ERP traces (Figure 2) and results of successive analyses in 50-msec latency windows, the following time windows were chosen for statistical analysis:[2] 0–250 msec (P100–N100–P200),

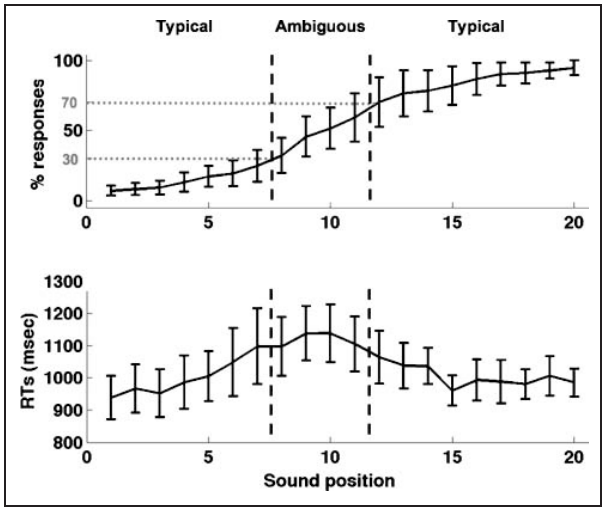


Figure 1. Categorization function (top) and mean RTs in millisecond (bottom) averaged across the 15 continua as a function of sound position (from 1 to 20). Standard deviations are indicated at each sound position. The categorization function represents the percentage of responses in the material category corresponding to the right extreme of the continuum. The vertical dotted gray lines delimit the zones of typical (extremes positions) and of ambiguous (intermediate positions) sounds.

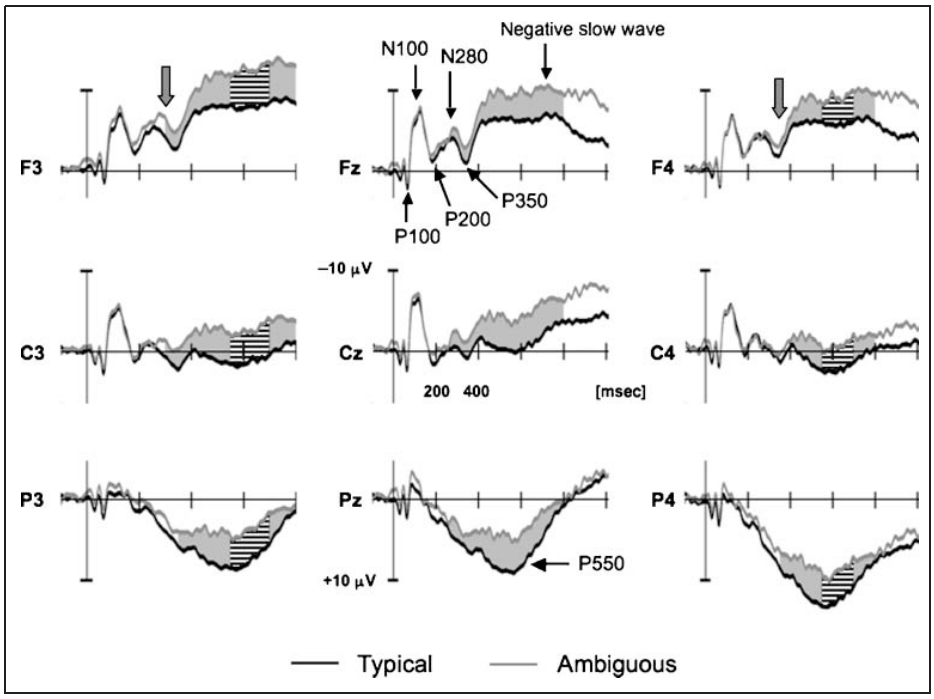


Figure 2. ERPs to typical (black line) and ambiguous sounds (gray line) at midline and at selected lateral electrodes (the most representative electrodes for each ROI). For this and following figures, the amplitude of the effects is represented on the ordinate (in μV; negativity is up). The time from sound onset is on the abscissa (in msec). The gray zones indicate the latency ranges in which differences between typical and ambiguous sounds were significant. These differences started earlier (vertical gray arrows) and were larger (striped zones) over the left than the right hemisphere.

250–400 msec (N280-P350), and 400–800 msec (negative slow wave [NSW] and P550).

Figure 2 shows the ERPs for midline and selected lateral electrodes (the most representative electrode for each ROI). Both typical and ambiguous sounds elicited similar P100, N100, and P200 components at midline and lateral electrodes (no significant effect in the 0- to 250-msec latency band). The ERPs in the two conditions then start to diverge with larger N280 and smaller P350 components in the 250- to 400-msec latency band for ambiguous (−0.89 μV) than for typical sounds (0.54 μV) at midline electrodes, $F(1, 21) = 4.19, p < .05$. Results of fine-grained analyses in successive 50-msec latency bands revealed that these differences started earlier over fronto-central regions of the left—from 300 msec at F3 ($p < .001$) and C3 ($p < .01$) electrodes; Type of Sounds × Hemispheres × ROI × Electrodes interaction: $F(4, 84) = 2.68, p < .05$—than of the right hemisphere—from 350 msec at F4 electrode ($p < .05$); Type of Sounds × Hemispheres × ROI × Electrodes interaction: $F(4, 84) = 2.32, p = .06$ (see Figure 2).

In the 400- to 800-msec latency range, the main effect of Type of Sounds was still significant: midline, $F(1, 21) = 23.60, p < .001$; lateral, $F(1, 21) = 17.91, p < .001$. Although typical sounds were associated with larger positivity (P550) than ambiguous sounds over parietal regions, ambiguous sounds elicited larger NSW than typical sounds over fronto-central regions: Type of Sounds × ROI interaction, $F(2, 42) = 4.11, p < .05$. These differences were larger over the left (3.4 μV) than the right (2.78 μV) hemisphere in the 550- to 700-msec latency band[2]: Type of sounds × Hemispheres interaction, $F(1, 21) = 4.60, p < .05$ (see striped zones in Figure 2).

Discussion

Analysis of behavioral data showed that sounds categorized within a material category by less than 70% of the participants (ambiguous sounds) were associated with slower RTs than sounds that were categorized within a category by more than 70% of the participants (typical sounds). This was found for each continuum. Thus, as hypothesized, ambiguous sounds were more difficult to categorize than typical sounds. This result is in line with previous findings in the literature showing slower RTs for nonmeaningful than for meaningful sounds (e.g., Cummings et al., 2006). The differences between typical and ambiguous sounds were smaller in the wood–metal and glass–metal continua than in the wood–glass continuum. This is interesting from an acoustic perspective because metal sounds typically present higher spectral complexity (related to the density and repartition of spectral components) than both wood and glass sounds that show closer sound properties. Thus, ambiguous sounds in wood–metal and glass–metal continua were easier to categorize than those in the wood–glass continuum and the ambiguity effect was smaller.

Electrophysiological data showed that ambiguous sounds elicited more negative ERPs (a negative component, N280, followed by an NSW) than typical sounds over fronto-central regions. By contrast, typical sounds elicited more positive ERPs (P350 and P550 components) than ambiguous sounds over frontal and parietal regions. These findings were expected on the basis of previous results in categorization tasks showing that the amplitude of the N200 component is larger and the amplitude of

the P300 component is smaller to stimuli that are more difficult to categorize (Donchin & Coles, 1988; Ritter et al., 1983; Duncan-Johnson & Donchin, 1982; Donchin, 1981; Kutas, McCarthy, & Donchin, 1977). Moreover, in line with the long duration of the RTs (around 1 sec), the long latency of the positive component (P550) is taken to reflect the difficulty of the categorization task (participants were required to categorize sounds in one of three possible categories) and the relatively long duration of the sounds (860 msec, on average, over the three categories). Thus, both behavioral and ERP data showed that we were able to create ambiguous sounds that were more difficult to categorize than typical sounds.

The differences between ambiguous and typical sounds started earlier over the left (300 msec) than the right (350 msec) hemisphere and were also larger over the left than right hemisphere in the 550- to 700-msec latency band (see striped zones in Figure 2). This scalp distribution is similar to the left-hemisphere distribution reported for sounds by Van Petten and Rheinfelder (1995). Moreover, as found by these authors, a long-lasting NSW developed over frontal sites that lasted for the entire recording period. Late NSW are typically interpreted as reflecting processes linked with the maintenance of stimuli in working memory, expectancy (Walter, Cooper, Aldridge, McCallum, & Winter, 1964), and attention (King & Kutas, 1995). In the present experiment, the NSW may indeed reflect expectancy processes because a row of XXXX followed sound offset, but it may also reflect sound duration processing (as the "sustained potential" reported by Alain, Schuler, & McDonald, 2002) and categorization difficulty because this fronto-central negativity was larger for ambiguous than for typical sounds. In particular, as it has been proposed for linguistic stimuli (see Kutas & Federmeier, 2000), this larger negativity may reflect the difficulty of accessing information from long-term memory.

Finally, it should be noted that no significant differences were found on the P100 and N100 components. These components are known to be sensitive to sound onset (e.g., attack time) and temporal envelope (for a review, see Kuriki, Kanda, & Hirata, 2006; Shahin, Roberts, Pantev, Trainor, & Ross, 2005; Shahin, Bosnyak, Trainor, & Roberts, 2003; Hyde, 1997). However, because differences in attack time between typical and ambiguous sounds were 0.1 msec, on average, they were consequently not perceptible as this value is below the temporal resolution of the human hearing system (Gordon, 1987), thereby explaining the lack of differences in the ERPs.

EXPERIMENT 2: CONCEPTUAL AND SEMANTIC PRIMING

Results of Experiment 1 showed that we were able to create typical and ambiguous sounds. The goal of Experiment 2 was to use these sounds in a priming design to address the three aims described in the introduction: (1) test for conceptual priming between pairs of nonlinguistic sounds, (2) use only one homogeneous class of sounds (impact sounds), and (3) directly compare conceptual priming for nonlinguistic stimuli on one side and for linguistic stimuli on the other side. To achieve these aims, it was important to use the same task with both types of sounds. Thus, participants were asked to decide whether the target belonged to the same or to a different category than the prime. For the linguistic sounds and based on the design used by Holcomb and Neville (1990), primes were always words, and targets were words, pseudowords, or nonwords (i.e., words played backward). To use similar design and experimental conditions with nonlinguistic sounds, primes were always impact sounds, and targets were typical impact sounds from the same category as the prime, ambiguous sounds and typical impact sounds from a different category than the prime.[3]

On the basis of previous results in the literature with linguistic stimuli (Holcomb & Neville, 1990; Bentin, McCarthy, & Wood, 1985) and results of Experiment 1 with nonlinguistic stimuli, we hypothesized that pseudowords and ambiguous sounds should be more difficult to categorize (i.e., higher error rates and slower RTs) than stimuli from the other two categories. Moreover, as reported by Holcomb and Neville (1990) and in previous studies (for a review, see Kutas, Van Petten, & Kluender, 2006), pseudowords should also elicit larger N400 than words. Holcomb and Neville (1990) argued that "Perhaps this was because their word-like characteristics also produce lexical activation, but because no complete match was achieved, the amount of activation produced was greater and more prolonged" (p. 306). More generally, this result has been taken to reflect the (unsuccessful) search for meaning of orthographically and phonologically legal constructions that nevertheless have no meaning (see Kutas & Federmeier, 2000). However, the N400 to pseudowords may also reflect their lower familiarity than words and their ambiguous nature: They are word-like at the orthographic and phonological levels but are not real words at the semantic level. In such case, ambiguous sounds that share acoustic properties with typical sounds of a material category but nevertheless are not typical exemplars of any categories may also elicit N400-like components. It was therefore of interest to determine whether ambiguous sounds would be processed as pseudowords and elicit N400-like components or would rather elicit increased N280 and NSW as found in Experiment 1. Finally, nonwords (i.e., words played backward) should elicit larger P300 components than words, as reported by Holcomb and Neville (1990). Indeed, although words played backward keep the main attributes of vocal sounds (i.e., the formantic structure of the spectrum due to the resonance of the vocal tract), they should readily be perceived as belonging to a different category than the word prime. Similarly, if typical sounds from a different category than the prime are easily categorized as such, they should also elicit larger P300 components than typical sounds from the same category than the prime.

Methods

Subjects

A total of 19 students (8 women and 11 men; 24 years old, on average) participated in this experiment that lasted for about 1 hour 30 min. None had participated in Experiment 1. They were right-handed, nonmusicians (no formal musical training), French native speakers with no known neurological disorders. They all gave written consent to participate in the experiment and were paid for their participation.

Stimuli

A total of 120 pairs of nonlinguistic stimuli were presented. Primes were always typical sounds from a given material category (i.e., wood, metal, or glass), and targets were either sounds from the same category as the prime (Same condition, 30 pairs), ambiguous sounds (Ambiguous condition, 60 pairs), or sounds from a different category than the prime (Different condition, 30 pairs). The number of ambiguous pairs was twice the number of pairs in the Same and Different conditions to balance the number of Yes and No responses. (On the basis of the results of Experiment 1, we expected participants to give as many Yes as No responses to ambiguous targets.) The averaged duration of nonlinguistic stimuli was 788 msec.

A total of 180 pairs of linguistic sounds were presented. Primes were always French spoken words and targets were spoken words (Same condition, 90 pairs), pseudowords (Ambiguous condition, 45 pairs), or nonwords (Different condition, 45 pairs). Word targets were bisyllabic nouns. Pseudowords were constructed by modifying one vowel from word targets (e.g., *boteau* from *bateau*). Nonwords were words played backward. The averaged duration of linguistic stimuli was 550 msec.

Procedure

Participants were asked to listen to each pair of stimuli and to determine whether the prime and the target belonged to the same category by pressing one of two response buttons. Nonlinguistic and linguistic stimuli were presented in two separate sessions 10 min apart with the linguistic session always presented after the nonlinguistic session. In each session, pairs of stimuli belonging to the three experimental conditions were randomly presented within three blocks of 40 trials for nonlinguistic pairs and three blocks of 60 trials for linguistic pairs (less nonlinguistic stimuli were presented within a block because sounds were longer in duration than linguistic stimuli).

To balance the number of Yes and No responses, each block of nonlinguistic stimuli comprised 10 same (yes), 20 ambiguous (yes/no), and 10 different pairs (no). Each block of linguistic stimuli comprised 30 Same (yes), 15 Ambiguous (no), and 15 Different (no) pairs. The order of block presentations within the nonlinguistic and linguistic sessions and the association between responses (Yes/No) and buttons (left/right) were balanced across participants.

For both nonlinguistic and linguistic pairs, targets followed prime offset with a 20-msec interstimulus interval. A row of XXXX was presented on the screen 2000 msec after target onset for 2000 msec to give participants time to blink. The prime of the next pair was then presented after a 1000-msec delay.

Recording ERPs

EEG was continuously recorded using the same procedure as in Experiment 1 and later segmented in single trials of 2200 msec starting 200 msec before target onset. Data were analyzed using the Brain Vision Analyzer software (Brain Products, Munich).

Nonlinguistic Sounds

Results

Behavioral data. For ambiguous sounds, there are no correct or incorrect responses because they can be associated to yes or to no responses. Thus, on the basis of the participants' responses, ANOVAs included Category as a factor with four conditions: Same, Ambiguous/Yes, Ambiguous/No, and Different targets. Results revealed a main effect of Category, $F(3, 54) = 111.71$, $p < .001$: Same and Different targets were associated with low error rates (6% and 4%, respectively) and did not differ from each other ($p = .92$). They differed from Ambiguous/Yes (46%; $p < .001$) and Ambiguous/No targets (53%; $p < .001$) that did not differ from each other ($p = .15$). Mean RTs were not significantly different ($p = .09$: 917 msec for Same, 900 msec for Ambiguous/Yes, 863 msec for Ambiguous/No, and 892 msec for Different targets).

Electrophysiological data. Two separate ANOVAs were conducted for midline and lateral electrodes. Category (Same, Ambiguous,[4] Different) and Electrodes (Fz, Cz, Pz) were included as factors for midline analyses. Category, Hemispheres (left vs. right), ROIs (fronto-central R1, centro-temporal R2, and centro-parietal R3), and Electrodes (3 for each ROI: [F7, F3, FC1]/[F8, F4, FC2]; [FC5, T7, C3]/[FC6, T8, C4]; and [CP1, CP5, P3]/[CP6, CP2, P4]) were included for lateral analyses. On the basis of visual inspection and results of successive analyses in 50-msec latency windows, time windows chosen for the statistical analysis were 0–150, 150–350, 350–450, 450–550, and 550–700 msec. For Same and Different targets, only correct responses were taken into account. Results are reported in Table 1.

Figure 3 (top) illustrates ERPs to nonlinguistic targets. In all conditions, sounds elicited P100, N100, P200, and

Table 1. Nonlinguistic Targets

(I)	*Factors*	*df*	*0–150 msec*	*150–350 msec*	*350–450 msec*	*450–550 msec*	*550–700 msec*
Midline	C	2,36	–	–	14.92***	17.13***	12.26***
	C × E	4,72	–	2.60*	2.50*	–	–
Lateral	C	2,36	–	–	15.70***	19.04***	12.10***
	C × ROI	4,72	–	–	2.44*	2.66*	4.89**

(II)	*150–350 msec*			*350–450 msec*		
C × E	*Fz*	*Cz*	*Pz*	*Fz*	*Cz*	*Pz*
A − S	–	–	–	−1.74**	−2.24	−2.52
A − D	−1.42*	–	–	−4.23	−3.75	−3.05
S − D	−1.9	–	–	−2.49	−1.51*	–

(III)	*350–450 msec*			*450–550 msec*			*550–700 msec*		
C × ROI	*R1*	*R2*	*R3*	*R1*	*R2*	*R3*	*R1*	*R2*	*R3*
A − S	−1.57	−1.8	−2.55	−1.54	−1.86	−2.37	−1.12**	−1.43	−1.44
A − D	−2.82	−2.51	−3.04	−3.17	−2.97	−4.09	−2.07	−2.04	−3.4
S − D	−1.25	–	–	−1.63	−1.11**	−1.72	−0.95*	–	−1.96

(I) *F* statistics for the main effect of Category (C), for the Category by Electrodes (C × E) interaction, and for the Category by ROI (C × ROI) interaction in the latency ranges chosen for analyses. (II and III) Mean amplitude differences (in μV) between Same (S), Ambiguous (A), and Different (D) conditions when the C × E and C × ROI interactions were significant. The reported difference values were always significant at $p < .001$ (results of post hoc tests) except when indicated by $p < .05$ or $p < .01$.

*$p < .05$.

**$p < .01$.

***$p < .001$.

N280 components followed by large negative components over fronto-central regions and P550 components over parietal regions. No significant effects were found in the latency range 0–150 msec (see Table 1-I). In the 150- to 350-msec latency range, the main effect of Category was not significant but the Category × Electrodes interaction was significant: Ambiguous and Same targets elicited larger N280 than Different targets at Fz (see Table 1-II). In the 350- to 450-msec latency range, the main effect of Category was significant with larger negativity to Ambiguous than to both Same and Different targets that did not differ from each other except over fronto-central region (see Table 1-II and 1-III). In the 450- to 550-msec latency range, the three conditions significantly differed from each other with the negativity being largest for Ambiguous, intermediate for Same, and the positivity largest for Different targets, with largest differences over centro-parietal regions (i.e., over R3 in Table 1-III). Finally, in the 550- to 700-msec latency range, at midline electrodes, Different targets elicited larger positivity (−0.01 μV) than Same (−2.05 μV) and Ambiguous targets (−3.28 μV) that did not differ from each other. By contrast, at lateral electrodes, the three conditions still differed significantly from each other with largest differences over the centro-parietal region.

Discussion

As hypothesized, the error rate for Same and Different targets was very low (6% and 4%, respectively), which shows that typical sounds were easily categorized as belonging to the same or to a different impact sound category than the prime. By contrast and as expected, on the basis of the results of Experiment 1, ambiguous targets were more difficult to categorize and were categorized as often as belonging to the same (46%) or to a different category (54%) from the prime. This clearly confirms the ambiguous nature of these sounds. The lack of effects on RTs may result from the relatively long duration of the prime sounds (788 msec, on average). Priming effects generally are short lived (Meyer & Schvaneveldt, 1971) and may consequently have vanished by the time the target sound was presented.

Regarding the ERPs, all targets elicited a P100–N100–P200 complex that, as expected (see Discussion of Experiment 1), did not differ between Same, Ambiguous,

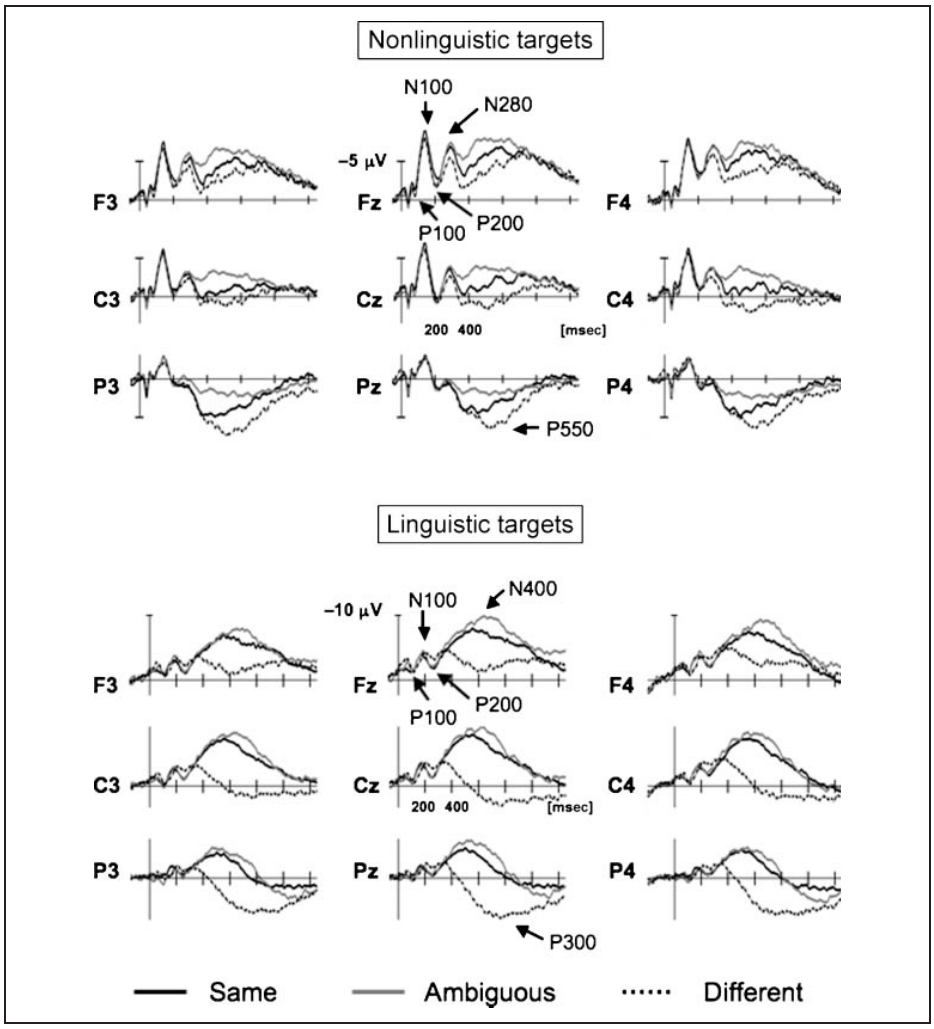


Figure 3. ERPs to Same (solid line), Ambiguous (gray line), and Different (dashed line) targets at midline and at selected lateral electrodes (the most representative electrodes for each ROI) for nonlinguistic (top) and linguistic (bottom) stimuli.

and Different target sounds. An N280 component was also elicited as in Experiment 1. Its amplitude was larger for Same and Ambiguous sounds than for Different sounds over frontal regions. However, the ERPs were morphologically different in Experiments 1 and 2. Although an NSW followed the N280 in Experiment 1 (and lasted until the end of the recording period), a temporally (between 350 and 700 msec) and spatially (fronto-central) localized negativity followed the N280 in Experiment 2. Fine-grained analyses allowed to specify the spatiotemporal dynamics of the effects. First, between 350 and 450 msec, the amplitude of this negative component was largest over fronto-central sites for Ambiguous targets, intermediate for Same targets, and smallest for Different targets. Then, between 450 and 550 msec, typical sounds from a different category than the prime elicited large P300 components over parietal sites thereby reflecting the fact that they were easily categorized as different (4% errors; Holcomb & Neville, 1990; Kutas et al., 1977). Because the same stimuli were used in Experiments 1 and 2, these differences are clearly linked with the task at hand (i.e., in Experiment 1, isolated impact sounds were to be categorized in one of three categories, whereas in Experiment 2, target sounds were compared with a prime). Thus, and as typically shown by fMRI data, these results demonstrate the strong influence of task demands on stimulus processing (e.g., Thierry, Giraud, & Price, 2003).

In Experiment 2, we used a priming design to be able to compare results with previous ones in the literature, and we presented two sounds and no words to reduce the use of linguistic strategies that, as described in the introduction, may have influenced previous results (Orgs et al., 2006, 2007, 2008; Cummings et al., 2006; Plante et al., 2000; Van Petten & Rheinfelder, 1995; and, to a lesser extent, Schön et al., 2009). The finding that a negative component developed in the 350- to 700-msec latency band with largest amplitude to Ambiguous sounds is in line with these previous studies and shows that conceptual priming can occur within sound–sound pairs. Moreover, this result was found when using the homogeneous class of impact sounds. However, before considering the implications of these results for conceptual priming, it is important to

examine results obtained for linguistic targets preceded by linguistic primes.

Linguistic Sounds

Results

Behavioral data. The main effect of Category (word [W], pseudoword [PW], nonword [NW]; within-subject factor) was significant, $F(2, 36) = 48.15$, $p < .001$: The error rate was higher for PW (13%) than for W (4.9%; $p < .001$) and NW (2.3%; $p < .001$). RTs were not significantly different ($p = .61$) for PW (1067 msec), W (1057 msec), and NW (1054 msec).

Electrophysiological data. Similar ANOVAs were conducted as for nonlinguistic sounds. Statistical analysis was conducted in the 0–150, 150–350, 350–600, 600–750, and 750–1100 msec latency ranges. Only correct responses were taken into account. Results of statistical analyses are reported in Table 2.

Figure 3 (bottom) illustrates ERPs to linguistic targets. No significant differences were found in the latency ranges 0–150 and 150–350 msec either at midline or at lateral

Table 2. Linguistic Targets

(I)	*Factors*	*df*	*0–150 msec*	*150–350 msec*	*350–600 msec*	*600–750 msec*	*750–1100 msec*
Midline	C	2,36	–	–	41.32***	90.52***	40.81***
	C × E	4,72	–	–	5.00**	8.97***	11.24***
Lateral	C	2,36	–	–	36.71***	85.53***	54.50***
	C × ROI	4,72	–	–	4.80**	7.41***	7.51***
	C × ROI × H	4,72	–	–	4.93**	2.87*	3.16*

(II)	*350–600 msec*			*600–750 msec*			*750–1100 msec*		
C × E	*Fz*	*Cz*	*Pz*	*Fz*	*Cz*	*Pz*	*Fz*	*Cz*	*Pz*
P − W	–	–	–	−2.57	−2.1	−2.22	−1.22*	–	–
P − N	−4.48	−6.33	−5.68	−7.79	−10.47	−9.58	−3.34	−5.61	−4.61
W − N	−3.53	−5.67	−4.62	−5.22	−8.37	−7.36	−2.12	−5.4	−4.6

(III)	*350–600 msec*			*600–750 msec*			*750–1100 msec*		
C × ROI	*R1*	*R2*	*R3*	*R1*	*R2*	*R3*	*R1*	*R2*	*R3*
P − W	–	–	–	−2.03	−1.57	−1.45	−0.95*	–	–
P − N	−3.03	−4.07	−4.37	−6.08	−7.11	−7.39	−2.58	−3.68	−3.47
W − N	−2.49	−3.65	−3.91	−4.05	−5.54	−5.94	−1.63	−3.25	−3.71

(IV)	*350–600 msec*		*600–750 msec*		*750–1100 msec*	
C × ROI × H	*L*	*R*	*L*	*R*	*L*	*R*
R1	–	−0.68*	−1.72	−2.33	–	−1.43
R2	–	–	−1.52	−1.61	–	–
R3	–	–	−1.61	−1.29	–	–

(I) *F* statistics for the main effect of Category (C), for the Category by Electrodes (C × E) interaction, for the Category by Regions of Interest (C × ROI) interaction, and for the Category by Regions of Interest by Hemispheres (C × ROI × H) interactions in the latency ranges of interest. (II and III) Mean amplitude differences (in μV) between Words (W), Pseudowords (P), and Nonwords (N) conditions for C × E and C × ROI interactions when effects were significant. The reported difference values were always significant at $p < .001$ (results of post hoc tests) except when indicated by $p < .05$. (IV) Mean amplitude differences P − W (in μV) for C × ROI × H interaction. The reported difference values were always significant at $p < .001$ (results of post hoc tests) except when indicated by $p < .05$.

*$p < .05$.

**$p < .01$.

***$p < .001$.

electrodes, but a main effect of Category was found in the 350–600, 600–750, and 750–1100 msec latency ranges at both midline and lateral electrodes (see Table 2-I). In these three latency ranges, NW always elicited larger positivity than both PW and W with largest differences at Cz and over centro-parietal regions (Table 2-II and 2-III). In addition, between 600 and 1100 msec, PW elicited larger negativity than W over right fronto-central regions (Table 2-IV).

Discussion

Behavioral data, showing higher error rate for PW than for both W and NW, are in line with previous results (e.g., Holcomb & Neville, 1990; Bentin et al., 1985). However, no effect was found on RTs, which again may reflect the relatively long duration of stimuli and of RTs (over 1 sec, on average) together with short-lived priming effects (Meyer & Schvaneveldt, 1971). As expected, on the basis of Holcomb and Neville's (1990) results, PW produced larger N400 components than W over anterior sites. Moreover, this N400 effect was larger over the right than the left hemisphere. This "paradoxical lateralization" (Plante et al., 2000, p. 1680) is consistent with previous results showing right-greater-than-left asymmetry of the N400 effect (Kutas et al., 1988; Kutas & Hillyard, 1982). Finally, the rather long latency of this N400 effect is also consistent with the results of Holcomb and Neville (1990), showing that the N400 effect starts earlier and lasts longer in the auditory than in the visual modality. It may also reflect the difficulty to categorize PW that were very similar to words (they were constructed by replacing only one vowel from an existing word). By contrast, NW (i.e., words played backward) was easy to categorize as different from the prime words and elicited a large P300 component with a posterior scalp distribution (Holcomb & Neville, 1990).

Nonlinguistic versus Linguistic Sounds

Because the same design was used for both nonlinguistic and linguistic sounds within the same group of participants, conceptual and semantic priming effects were directly compared by including Stimulus (nonlinguistic vs. linguistic) as a factor.

ANOVAs were conducted in the 350- to 800-msec time window, where significant differences were found for both nonlinguistic and linguistic sounds. Results of statistical analyses are reported in Table 3. The main effect of Stimulus was significant: ERPs to linguistic stimuli were overall more negative than to nonlinguistic stimuli (Table 3-II). Moreover, the main effect of Category was significant with largest N400 to Ambiguous (ambiguous impact sounds and PW), intermediate to Same (same impact sounds and W), and largest positivity to Different targets (different impact sounds and NW) (Table 3-III). Finally, the Stimulus × Category interaction was significant. Although the difference between Ambiguous and Same targets was similar for both linguistic and nonlinguistic stimuli, the difference between Different and Same targets was significantly larger for linguistic than for nonlinguistic stimuli (Table 3-IV and Figure 4).

Table 3. Nonlinguistic versus Linguistic Targets

(I)	*Factors*	*df*	*F*
Midline	Stimulus	1,18	17.48***
	C	2,36	68.57***
	Stimulus × C	2,36	14.43***
Lateral	Stimulus	1,18	24.18***
	C	2,36	65.95***
	Stimulus × C	2,36	12.69***

(II)	*Midline*	*Lateral*
Nonlinguistic	−1.52	−0.42
Linguistic	−4.00	−2.98

(III)	*Midline*	*Lateral*
S	−3.51	−2.19
A	−4.95	−3.35
D	0.17	0.43

(IV)		*Midline*	*Lateral*
A − S	Nonlinguistic	1.43	1.31
	Linguistic	1.46	1.01
D − S	Nonlinguistic	1.80	1.17
	Linguistic	5.57	4.06

(I) *F* statistics for the main effect of Stimulus and Category (C) and for the Stimulus by Category (Stimulus × C) interaction in the 350- to 800-msec latency range. (II) Mean amplitude (in μV) of the main effect of Stimulus. (III) Mean amplitude (in μV) of the main effect of Category: Same (S), Ambiguous (A), and Different (D) conditions. (IV) Mean amplitude differences A − S and D − S (in μV) for Nonlinguistic and Linguistic stimuli.

***$p < .001$.

GENERAL DISCUSSION

Results of the general ANOVA highlighted clear similarities between conceptual priming for nonlinguistic and linguistic sounds. In both cases, behavioral data showed higher error rates in the Ambiguous than in the Same and Different conditions with no effects on RTs. This ambiguity effect most likely reflects the difficulty to correctly categorize Ambiguous targets as different because they are similar to the prime (e.g., orthographic and phonologic similarity for PW and acoustic proximity for impact sounds). Several studies using priming designs showed higher error rates

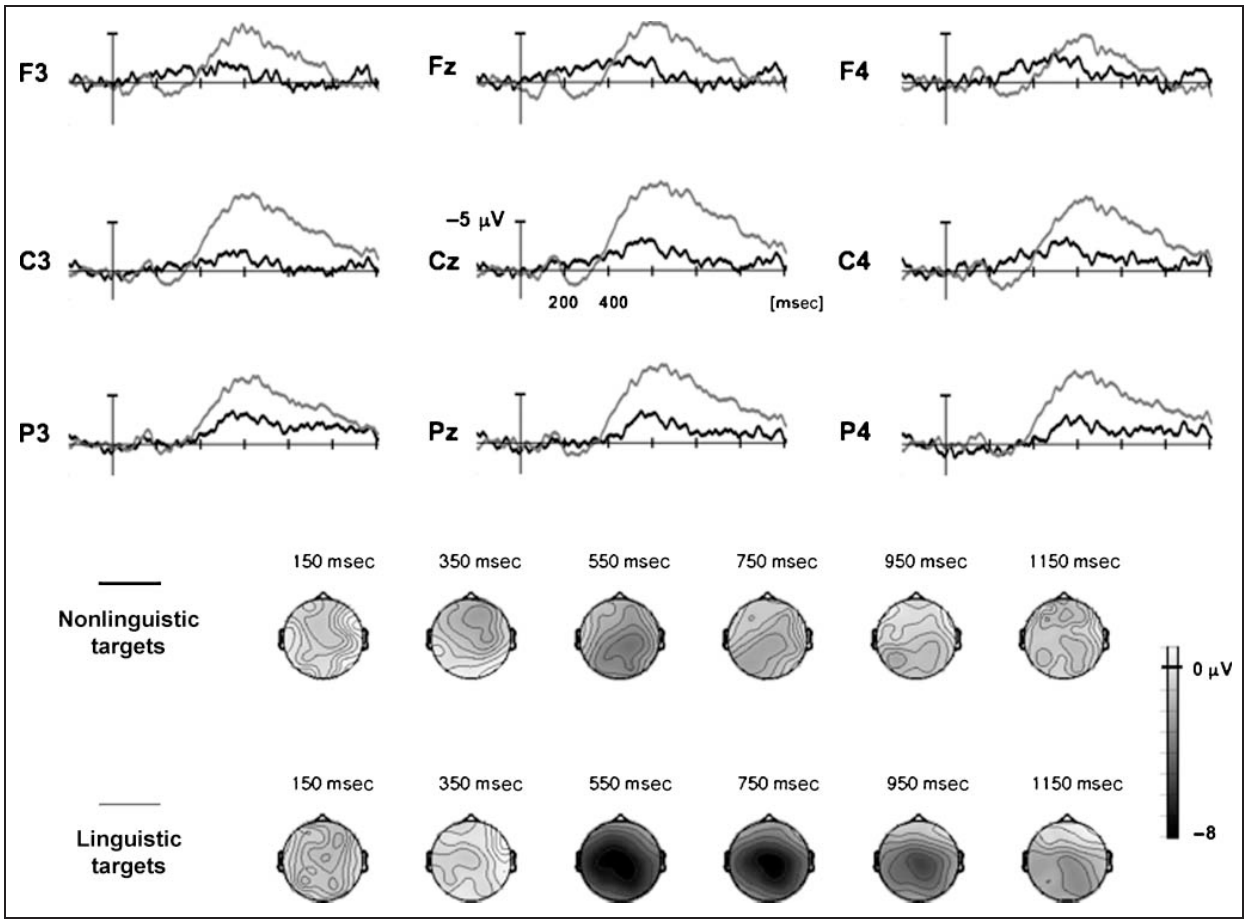


Figure 4. Same-minus-Different Difference Waves. ERPs to nonlinguistic (black line) and linguistic (gray line) targets at midline and at selected lateral electrodes (the most representative electrodes for each ROI). Temporal dynamics of the scalp distribution of the effects from 150 to 1150 msec for nonlinguistic and linguistic targets.

for PW than for W (e.g., Holcomb & Neville, 1990) as well as for unrelated than for related words (e.g., Bentin et al., 1985; Boddy, 1981). By contrast, results differ in some studies using nonlinguistic sounds. For instance, Orgs et al. (2006, 2008) found higher error rates in related compared with unrelated pairs. They explained this result by the greater ambiguity of environmental sounds due to causal uncertainties that influence their labeling.

Most interestingly, analyses of the ERPs revealed similar modulation of the late components elicited by nonlinguistic and linguistic sounds: largest negativity for Ambiguous, intermediate for Same, and largest positivity for Different targets. These differences emerged with similar onset latencies in both cases (i.e., at 350 msec after target onset). Importantly, the Stimulus × Category interaction was significant: differences between Same and Different targets were larger for linguistic than for nonlinguistic sounds (see Figure 4). Because linguistic Different targets were words played backward, they were unfamiliar stimuli. Therefore, they were probably more surprising than nonlinguistic Different targets that were typical impact sounds and consequently more familiar but still different from the prime. By contrast, the priming effect for Ambiguous stimuli was similar in the linguistic and in the nonlinguistic conditions (i.e., the difference between Ambiguous and Same categories was not significantly different, either in amplitude or in scalp distribution for nonlinguistic and for linguistic stimuli).

However, results of separate ANOVAs nevertheless revealed that the spatiotemporal dynamics of the ambiguity effect was somewhat different for nonlinguistic and linguistic sounds, with an earlier onset for ambiguous impact sounds than for PW and a slight predominance over right frontal sites for PW. As noted in the introduction, although priming studies using environmental sounds have reported ERP effects that closely resemble the verbal N400 effect, they also showed differences in scalp distribution. As found here, priming effects were larger over the right hemisphere for words and over the left hemisphere for environmental sounds (Plante et al., 2000; Van Petten & Rheinfelder, 1995). By contrast, Orgs et al. (2006, 2008) found no interhemispheric differences but larger priming

effects for sounds over posterior than anterior sites, and Cummings et al. (2006) found larger differences over anterior than posterior regions (as found here). Thus, the scalp topography seems somewhat variable between experiments, which most likely reflects differences in the acoustic properties of the stimuli and in task demands.

This conclusion is in line with results in the fMRI literature on verbal and environmental sounds showing mixed evidence in favor of the similarity of conceptual priming with nonlinguistic and linguistic sounds. For instance, although both spoken words and environmental sounds activate bilateral temporal regions (Giraud & Price, 2001; Humphries, Willard, Buchsbaum, & Hickok, 2001), Thierry et al. (2003) have demonstrated larger activation of the left anterior and posterior temporal areas for spoken words and larger activation of the right posterior superior temporal areas for environmental sounds. These between-experiments differences were taken to reflect differences in the task semantic requirements. Recently, Steinbeis and Koelsch (2008) provided evidence for both similar and different neural activations related to the processing of meaning in music and speech.

Taken together, our results are in line with previous literature (Schön et al., 2009; Daltrozzo & Schön, 2009; Orgs et al., 2006, 2007, 2008; Cummings et al., 2006; Plante et al., 2000; Van Petten & Rheinfelder, 1995) and argue in favor of the similarity of conceptual priming for nonlinguistic and linguistic sounds. Interestingly, the present results extend previous ones in several aspects. Most importantly, previous results were problematic in that words were always included in the design. As a consequence, the reported conceptual priming effects were possibly due to a linguistic strategy of generating words when listening to sounds. Although we also used linguistic stimuli to be able to compare priming effects within subjects, they were always presented in a separate session. The finding of N400-like components in a sound–sound design, as used in Experiment 2, shows that linguistic mediation is not necessary for an N400-like component to be elicited. Thus, this component may reflect a search for meaning that is not restricted to linguistic meaning. This interpretation is in agreement with the idea that variations in N400 amplitude are related to the "ease or difficulty of retrieving stored conceptual knowledge associated with a word or other meaningful stimuli" (Kutas et al., 2006, p. 10). Moreover, although two conditions (related vs. unrelated) were used in most previous studies, we used three conditions (Same vs. Ambiguous vs. Different) to more closely examine conceptual priming effects. In line with early studies of category membership effects (Ritter et al., 1983; Vaughan, Sherif, O'Sullivan, Herrmann, & Weldon, 1982; Boddy, 1981; Boddy & Weinberg, 1981), stimuli that clearly did not belong to the prime category elicited late positivity (P300 components), whereas stimuli that were ambiguous elicited late negativity (N400-like components) compared with stimuli that belonged to the prime category. Most importantly for our purposes, we were able to demonstrate similar relationships between categories for both nonlinguistic and linguistic target sounds.

Conclusion

These results add interesting information to the vast and still largely unexplored domain of the semiotics of sounds. Other experiments using different tasks and stimuli are needed to further explore the similarities and differences in conceptual priming for nonlinguistic and linguistic sounds. However, by using a homogeneous class of environmental sounds (impact sounds), by varying the relationship between prime and target sounds, and by comparing conceptual priming for nonlinguistic and for linguistic sounds within the same participants, we were able to make one step further and to show that conceptual priming develops in a sound–sound design without words and, consequently, that conceptual priming can develop without (or with reduced) linguistic mediation.

Acknowledgments

This research was supported by a grant from the Human Frontier Science Program (HFSP #RGP0053) to Mireille Besson (PI) and from the French National Research Agency (ANR, JC05-41996, "senSons") to Sølvi Ystad (PI).

Reprint requests should be sent to Mitsuko Aramaki, CNRS—Institut de Neurosciences Cognitives de la Méditerranée, 31, Chemin Joseph Aiguier, 13402 Marseille Cedex 20, France, or via e-mail: Mitsuko.Aramaki@incm.cnrs-mrs.fr.

Notes

1. The Continua factor was not taken into account to keep enough trials in each condition.
2. Fine-grained analyses were computed as separated ANOVAs (that included the same factors as described in the Results section) in successive 50-msec latency windows from 0 to 800 msec after sound onset. Then, the 50-msec latency windows within which statistically similar effects were found were grouped together between 0–250, 250–400, and 400–800 msec or more specifically between 550 and 700 msec, and an ANOVA was conducted in each latency band.
3. To increase the similarities between the Different conditions for nonlinguistic and linguistic sounds, we also considered the possibility of playing impact sounds backward as was done for the words. However, although such sounds conserve the spectral characteristics of the original sound (i.e., acoustic cues characterizing the material category), they do no longer sound as impact sounds (i.e., the perception of impact disappears). They are therefore ambiguous and difficult to categorize. Because it was important to equate task difficulty for nonlinguistic and linguistic sounds (the words played backward are easy to categorize as different from the prime) and because words played backward keep the main attributes of vocal sounds, we decided to use typical sounds from another material category that are not ambiguous and easy to categorize as different.
4. On the basis of the results of behavioral data showing no differences between ambiguous targets associated with yes and no responses, we averaged ERPs in these two categories together to increase the signal to noise ratio. Moreover, no differences were found when ERPs to Ambiguous/Yes and Ambiguous/No were averaged separately.

REFERENCES

Alain, C., Schuler, B. M., & McDonald, K. L. (2002). Neural activity associated with distinguishing concurrent auditory objects. *Journal of the Acoustical Society of America, 111,* 990–995.

Aramaki, M., Baillères, H., Brancheriau, L., Kronland-Martinet, R., & Ystad, S. (2007). Sound quality assessment of wood for xylophone bars. *Journal of the Acoustical Society of America, 121,* 2407–2420.

Aramaki, M., & Kronland-Martinet, R. (2006). Analysis–synthesis of impact sounds by real-time dynamic filtering. *IEEE Transactions on Audio, Speech, and Language Processing, 14,* 695–705.

Ballas, J. A. (1993). Common factors in the identification of an assortment of brief everyday sounds. *Journal of Experimental Psychology: Human Perception and Performance, 19,* 250–267.

Ballas, J. A., & Howard, J. H., Jr. (1987). Interpreting the language of environmental sounds. *Environment and Behavior, 19,* 91–114.

Bentin, S., McCarthy, G., & Wood, C. C. (1985). Event-related potentials, lexical decision, and semantic priming. *Electroencephalography and Clinical Neurophysiology, 60,* 343–355.

Boddy, J. (1981). Evoked potentials and the dynamics of language processing. *Biological Psychology, 13,* 125–140.

Boddy, J., & Weinberg, H. (1981). Brain potentials, perceptual mechanisms and semantic categorization. *Biological Psychology, 12,* 43–61.

Cummings, A., Ceponiene, R., Koyama, A., Saygin, A. P., Townsend, J., & Dick, F. (2006). Auditory semantic networks for words and natural sounds. *Brain Research, 1115,* 92–107.

Daltrozzo, J., & Schön, D. (2009). Conceptual processing in music as revealed by N400 effects on words and musical targets. *Journal of Cognitive Neuroscience, 21,* 1882–1892.

de Saussure, F. (1916). *Cours de linguistique générale*. Paris: Payot.

Donchin, E. (1981). Surprise!…surprise? *Psychophysiology, 18,* 493–513.

Donchin, E., & Coles, M. G. H. (1988). Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences, 11,* 357–374.

Duncan-Johnson, C. C., & Donchin, E. (1982). The P300 component of the event-related brain potential as an index of information processing. *Biological Psychology, 14,* 1–52.

Frey, A., Marie, C., Prod'Homme, L., Timsit-Berthier, M., Schön, D., & Besson, M. (2009). Temporal semiotic units as minimal meaningful units in music? An electrophysiological approach. *Music Perception, 26,* 247–256.

Giraud, A. L., & Price, C. J. (2001). The constraints functional neuroimaging places on classical models of auditory word processing. *Journal of Cognitive Neuroscience, 13,* 754–765.

Gordon, J. W. (1987). The perceptual attack time of musical tones. *Journal of the Acoustical Society of America, 82,* 88–105.

Holcomb, P. J., & Neville, H. J. (1990). Auditory and visual semantic priming in lexical decision: A comparison using event-related brain potentials. *Language and Cognitive Processes, 5,* 281–312.

Humphries, C., Willard, K., Buchsbaum, B., & Hickok, G. (2001). Role of anterior temporal cortex in auditory sentence comprehension: An fMRI study. *NeuroReport, 12,* 1749–1752.

Hyde, M. (1997). The N1 response and its applications. *Audiology & Neuro-otology, 2,* 281–307.

Jasper, H. H. (1958). The ten–twenty electrode system of the international federation. *Electroencephalography and Clinical Neurophysiology, 10,* 371–375.

King, J., & Kutas, M. (1995). Who did what and when? Using word- and clause-level ERPs to monitor working memory usage in reading. *Journal of Cognitive Neuroscience, 7,* 376–395.

Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T., & Friederici, A. (2004). Music, language and meaning: Brain signatures of semantic processing. *Nature Neuroscience, 7,* 302–307.

Kronland-Martinet, R., Guillemain, P., & Ystad, S. (1997). Modelling of natural sounds by time-frequency and wavelet representations. *Organised Sound, 2,* 179–191.

Kuriki, S., Kanda, S., & Hirata, Y. (2006). Effects of musical experience on different components of MEG responses elicited by sequential piano-tones and chords. *Journal of Neuroscience, 26,* 4046–4053.

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension, Language comprehension and the N400. *Trends in Cognitive Sciences, 4,* 463–470.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207,* 203–204.

Kutas, M., & Hillyard, S. A. (1982). The lateral distribution of event-related potentials during sentence processing. *Neuropsychologia, 20,* 579–590.

Kutas, M., McCarthy, G., & Donchin, E. (1977). Augmenting mental chronometry: The P300 as a measure of stimulus evaluation time. *Science, 197,* 792–795.

Kutas, M., Van Petten, C., & Besson, M. (1988). Event-related potential asymmetries during the reading of sentences. *Electroencephalography and Clinical Neurophysiology, 69,* 218–233.

Kutas, M., Van Petten, C., & Kluender, R. (2006). Handbook of psycholinguistics. In M. A. Gernsbacher & M. Traxler (Eds.), *Psycholinguistics electrified II (1994–2005)* (2nd ed., pp. 659–724). New York: Elsevier Press.

Lebrun, N., Clochon, P., Etévenon, P., Lambert, J., Baron, J. C., & Eustache, F. (2001). An ERD mapping study of the neurocognitive processes involved in the perceptual and semantic analysis of environmental sounds and words. *Cognitive Brain Research, 11,* 235–248.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology, 90,* 227–234.

Orgs, G., Lange, K., Dombrowski, J., & Heil, M. (2006). Conceptual priming for environmental sounds and words: An ERP study. *Brain and Cognition, 62,* 267–272.

Orgs, G., Lange, K., Dombrowski, J., & Heil, M. (2007). Is conceptual priming for environmental sounds obligatory? *International Journal of Psychophysiology, 65,* 162–166.

Orgs, G., Lange, K., Dombrowski, J. H., & Heil, M. (2008). N400-effects to task-irrelevant environmental sounds: Further evidence for obligatory conceptual processing. *Neuroscience Letters, 436,* 133–137.

Plante, E., Van Petten, C., & Senkfor, A. J. (2000). Electrophysiological dissociation between verbal and nonverbal semantic processing in learning disabled adults. *Neuropsychologia, 38,* 1669–1684.

Ritter, W., Simson, R., & Vaughan, H. G. (1983). Event-related potential correlates of two stages of information processing in physical and semantic discrimination tasks. *Psychophysiology, 20,* 168–179.

Ritter, W., Simson, R., Vaughan, H. G., & Friedman, D. (1979). A brain event related to the making of sensory discrimination. *Science, 203,* 1358–1361.

Schön, D., Ystad, S., Kronland-Martinet, R., & Besson, M. (2009). The evocative power of sounds: Conceptual priming between words and nonverbal sounds. *Journal of Cognitive Neuroscience, 22,* 1026–1035.

Shahin, A., Bosnyak, D. J., Trainor, L. J., & Roberts, L. E. (2003). Enhancement of neuroplastic P2 and N1c auditory evoked potentials in musicians. *Journal of Neuroscience, 23,* 5545–5552.

Shahin, A., Roberts, L. E., Pantev, C., Trainor, L. J., & Ross, B. (2005). Modulation of P2 auditory-evoked responses by the spectral complexity of musical sounds. *NeuroReport, 16,* 1781–1785.

Simson, R., Vaughan, H. G., & Ritter, W. (1977). The scalp topography of potentials in auditory and visual discrimination tasks. *Electroencephalography and Clinical Neurophysiology, 42,* 528–535.

Steinbeis, N., & Koelsch, S. (2008). Comparing the processing of music and language meaning using EEG and fMRI provides evidence for similar and distinct neural representations. *PLoS ONE, 3,* e2226. doi:10.1371/journal.pone.0002226.

Thierry, G., Giraud, A.-L., & Price, C. (2003). Hemispheric dissociation in access to the human semantic system. *Neuron, 38,* 499–506.

Van Petten, C., & Rheinfelder, H. (1995). Conceptual relationships between spoken words and environmental sounds: Event-related brain potential measures. *Neuropsychologia, 33,* 485–508.

Vaughan, J., Sherif, K., O'Sullivan, R. L., Herrmann, D. J., & Weldon, D. A. (1982). Cortical evoked responses to synonyms and antonyms. *Memory and Cognition, 10,* 225–231.

Walter, W. G., Cooper, R., Aldridge, V. J., McCallum, W. C., & Winter, A. L. (1964). Contingent negative variation: An electrical sign of sensorimotor association and expectancy in the human brain. *Nature, 230,* 380–384.

Journal of Abnormal Psychology
2012, Vol. 121, No. 3, 628–640

© 2011 American Psychological Association
0021-843X/11/$12.00 DOI: 10.1037/a0026411

Toward an Exploration of Feeling of Strangeness in Schizophrenia: Perspectives on Acousmatic and Everyday Listening

Jean-Arthur Micoulaud-Franchi
Centre Hospitalier Universitaire de Sainte-Marguerite, Marseille, France, and Institut de Neurosciences Cognitives de la Méditerranée, Marseille, France

Mitsuko Aramaki
Institut de Neurosciences Cognitives de la Méditerranée, Marseille, France

Adrien Merer
Laboratoire de Mécanique et d'Acoustique, Marseille, France

Michel Cermolacce
Centre Hospitalier Universitaire de Sainte-Marguerite, Marseille, France, and Institut de Neurosciences Cognitives de la Méditerranée, Marseille, France

Solvi Ystad and Richard Kronland-Martinet
Laboratoire de Mécanique et d'Acoustique, Marseille, France

Jean Naudin and Jean Vion-Dury
Centre Hospitalier Universitaire de Sainte-Marguerite, Marseille, France, and Institut de Neurosciences Cognitives de la Méditerranée, Marseille, France

The aim of this study was to investigate abnormal perceptual experiences in schizophrenia, in particular the feeling of strangeness, which is commonly found in patients' self-reports. The experimental design included auditory complex stimuli within 2 theoretical frameworks based on "sensory gating deficit" and "aberrant salience," inspired from conventional perceptual scales. A specific sound corpus was designed with environmental (meaningful) and abstract (meaningless) sounds. The authors compared sound evaluations on 3 perceptual dimensions (*bizarre*, *familiar*, and *invasive*) and 2 emotional dimensions (*frightening* and *reassuring*) between 20 patients with schizophrenia (SCZ) and 20 control participants (CTL). The perceptual judgment was rated on independent linear scales for each sound. In addition, the conditioning-testing P50 paradigm was conducted on 10 SCZ and 10 CTL. Both behavioral and electrophysiological data confirmed the authors' expectations according to the 2 previous theoretical frameworks and showed that abnormal perceptual experiences in SCZ consisted of perceiving meaningful sounds in a distorted manner and as flooding/inundating but also in perceiving meaningless sounds as things that become meaningful by assigning them some significance. In addition, the use of independent scales to each perceptual dimension highlighted an unexpected ambivalence on familiarity and bizarreness in SCZ compatible with the explanation of semantic process impairment. The authors further suggested that this ambivalence might be due to a conflicting coactivation of 2 types of listening, that is, every day and musical (or acousmatic) listening.

Keywords: schizophrenia, perceptual abnormalities, P50, sensory gating, aberrant salience

Abnormal perception of external stimuli constitutes an important phenomenological feature of subjective experience in schizophrenia (Uhlhaas & Mishara, 2007). In the retrospective study of Cutting and Dunne (1989) that investigated subjective experience in patients, the most reliable feature that differentiated schizophrenia from depression was based on "perceptual anomalies." Many scales have explored and confirmed the presence of abnormal perceptual experiences in schizophrenia, such as the Structured Interview for Assessing Perceptual Anomalies (SIAPA), the Aberrant Salience Inventory (ASI), or the Cardiff Anomalous Perceptions Scale (CAPS) (Bell, Halligan, & Ellis, 2006; Bunney et al., 1999; Cicero, Kerns, & McCarthy, 2010). As shown by these scales, the common feature in the schizophrenic perceptual experience may be related to a "feeling of strangeness" or a "sensation of unusualness" (Bell et al., 2006; Cermolacce, Sass, & Parnas, 2010). This feeling is revealed by the patients' difficulties in experiencing a familiar, veridical and shareable world, which could induce a deep handicap in everyday situation. Apart from some analyses inspired by philosophical phenomenology (Blankenburg & Mishara, 2001; Cermolacce et al., 2010; Sass, 2001; Stanghellini, 2000; Wiggins, Schwartz, & Naudin, 2001), the

This article was published Online First December 12, 2011.

Jean-Arthur Micoulaud-Franchi, Michel Cermolacce, Jean Naudin, and Jean Vion-Dury, Pôle de Psychiatrie "Solaris," Centre Hospitalier Universitaire de Sainte-Marguerite, Marseille, France, and INCM-CNRS UMR 6193, Institut de Neurosciences Cognitives de la Méditerranée, Marseille, France; Mitsuko Aramaki, INCM-CNRS UMR 6193, Institut de Neurosciences Cognitives de la Méditerranée, Marseille, France; Adrien Merer, Solvi Ystad, and Richard Kronland-Martinet, LMA-CNRS UPR 7051, Laboratoire de Mécanique et d'Acoustique, Marseille, France.

We thank Pr. Christophe Lançon and Pr Jean-Michel Azorin.

Correspondence concerning this article should be addressed to Jean-Arthur Micoulaud-Franchi, J.A. Pôle de Psychiatrie "Solaris," Centre Hospitalier Universitaire de Sainte-Marguerite, 270 Bd de Sainte-Marguerite, 13009 Marseille, France. E-mail: jarthur.micoulaud@gmail.com

feeling of strangeness remains an undefined and general concept (Bell et al., 2006) largely unexplored in an experimental paradigm (Cermolacce et al., 2010).

The principal aim of this study was to better investigate this feeling of strangeness, among more general altered perceptual experiences when auditory stimuli are presented to patients with schizophrenia and control participants. To interpret the behavioral and electrophysiological results of this study, two main perspectives afford us a theoretical framework, which are inspired from the perceptual scales previously cited: the hypothesis of "sensory gating deficit" (related to SIAPA; Bunney et al., 1999) and the "aberrant salience" hypothesis (related to ASI; Cicero et al., 2010). These two frameworks will be considered as complementary.

Abnormal Perceptual Experiences and Sensory Gating Deficit Hypothesis

In their now-classic phenomenological study based on interviews, McGhie and Chapman (1961) concluded that abnormal perceptual experiences of patients were related to a primary deficit in the selective and inhibitory functions of attention. Their finding might explain that patients generally feel as if they are being flooded by an overwhelming mass of sensory input combined with a heightened sensory perception, particularly in the auditory and visual modalities (McGhie & Chapman, 1961). Since this study, it has been suggested that alterations in the neurobiological process related to filtering stimuli, focusing attention, or sensory gating may explain abnormal perceptual experiences in schizophrenia (Andreasen et al., 1994). As a consequence, schizophrenia patients may be deficient in their ability to process the relevant information and to attribute coherent meaning to sensory inputs (Braff & Geyer, 1990).

The SIAPA was inspired by these studies and particularly by the McGhie and Chapman's pioneer work (Bunney et al., 1999). For the five sensory modalities, typical Likert items, based on self-reports from patients, were assessed on three dimensions: *hypersensitivity*, *inundation–flooding*, and *selective attention to external usual stimuli*, based on self-reports of patients. They reported a significantly greater prevalence of auditory and visual perceptual anomalies in patients with schizophrenia, when compared to control subjects (Bunney et al., 1999) and confirmed the results of McGhie and Chapman (1961).

An electrophysiological paradigm was proposed to better specify the sensory gating phenomenon. In the auditory modality, the event related potential (ERP) method was used to measure sensory gating in a click-paired-stimulus (S1-S2) or *conditioning-testing P50 paradigm* (Freedman, Adler, Waldo, Pachtman, & Franks, 1983). In healthy subjects, it was found that the amplitude of the P50 component, a positive ERP component occurring around 50 ms after stimulation onset, is smaller (down to half of the value) for the second stimulus (S2) than for the first stimulus (S1) of the pair (Adler et al., 1982; Freedman et al., 1987; Freedman et al., 1983). The P50 amplitude ratio between S2 and S1 commonly serves as a measure of the auditory gating and, by extension, of the sensory gating. It is well established that this ratio is greater for schizophrenia patients than for healthy subjects, underlying deficient neural sensory gating (de Wilde, Bour, Dingemans, Koelman, & Linszen, 2007; Patterson et al., 2008). It has been proposed that two patterns may contribute to the lack of decrease of this ratio in schizophrenia patients (Boutros & Belger, 1999; Brenner et al., 2009; Bunney et al., 1999). The first pattern is because the P50 amplitude is not reduced for stimulus S2. This gating out deficit suggests that patients might present a deficiency in the inhibitory mechanisms activated by a duplicated stimulus (S2) and consequently, in their ability to filter out redundant or irrelevant stimuli. The second pattern is because the P50 amplitude elicited by stimulus S1 is abnormally small. This gating in deficit suggests that patients might be deficient in their ability to encode and register new sensory inputs.

It was proposed that the sensory gating deficiency assessed by conditioning-testing P50 paradigm might be a good candidate for a neuronal substrate of abnormal perceptual experiences in schizophrenia further revealed by the SIAPA scale (Bunney et al., 1999; Freedman et al., 1987). However, in contrast to this assumption, Jin et al. (1998) did not find a relationship between abnormal experiences on auditory and visual items of the SIAPA and sensory gating deficits. Actually, the link between electrophysiological characterization of sensory gating anomalies by the P50 paradigm and phenomenological perceptual anomalies is still not well established and little investigated (Hetrick, Erickson, & Smith, 2010; Johannesen, Bodkins, O'Donnell, Shekhar, & Hetrick, 2008; Light & Braff, 2000). Indeed some contradictory results are found in the concerned literature, which could be explained by the difficulty in linking neurophysiological data to retrospective self-report questionnaire (Jin et al., 1998). Our study may bring some new data to better address this question.

Abnormal Perceptual Experiences and Aberrant Salience Hypothesis

Kapur (2003) suggested that abnormal perceptual experiences in schizophrenia could be explained by an aberrant salience of sensory inputs. In particular, Kapur hypothesized that an incorrect assignment of salience and significance to innocuous meaningless stimuli may constitute a central mechanism of schizophrenia (Kapur, 2003). This hypothesis was supported by patients' reports and by dysregulation of the mesolimbic dopamine system in schizophrenia patients (Heinz & Schlagenhauf, 2010).

Recently, the ASI aimed at investigating the abnormal perceptual experiences in relation with the aberrant salience hypothesis (Cicero et al., 2010). The ASI explores five factors according to 29 items, among them, the increased significance factor (evaluated by item like "Do certain trivial things ever suddenly seem epically important or significant to you?"), to which participants responded *yes* or *no*. Cicero et al. (2010) confirmed an increased attribution of meaning to external meaningless things in schizophrenia in line with Kapur's hypothesis (2003).

Our Study

The principal aim of this study was to explore the abnormal perceptual experiences in schizophrenia with the frameworks provided by the two hypotheses (sensory gating deficit and aberrant salience). Because abnormal perceptual experiences were found more frequently in auditory modality in schizophrenic patients (Bunney et al., 1999), the experimental material of this study consists in nonverbal complex sounds. We designed our sound corpus according to the two hypotheses previously presented, in a

complementary way, by collecting both *environmental sounds* and a specific class of sounds called *abstract* sounds. Environmental sounds are usual sounds that would be easily associated with a specific meaning from everyday life situations and thus be experienced as familiar. By contrast, abstract sounds are defined as unusual sounds and generally not encountered in the surrounding everyday world. They could not be easily associated with a physical sound source or a consensual meaning and be experienced as bizarre. By way of example, various terms that refer to abstract sounds can be found in the literature such as "strange" or "meaningless" (Solomon, 1958; see Merer, Ystad, Kronland-Martinet, & Aramaki, 2010, for a review). Abstract sounds were widely investigated by the electroacoustic music community, even if the term abstract was not directly used. In particular, electroacoustic music composers have developed specific recording and signal-processing techniques to avoid the clear recognition of the physical sources for musical applications. In 1966, the French composer and pioneer of "musique concrete," Pierre Schaeffer, who was both a musician and a researcher, introduced the concept of the so-called "acousmatic listening" in his book, *Traité des Objets Musicaux* (Schaeffer, 1966). Schaeffer defined acousmatic listening as the experience of listening to a sound without considering an originating cause. Abstract sounds may be considered as sounds that enhance acousmatic listening and inhibit the organization of auditory information in a coherent meaning (Merer et al., 2010; Schon, Ystad, Kronland-Martinet, & Besson, 2011).

Thus, based on the frameworks provided by the two previous hypotheses, abstract sounds (meaningless sounds) may be considered as an appropriate material to explore the aberrant salience hypothesis, in addition to environmental sounds (meaningful sounds) that may be considered as an appropriate material to explore the sensory gating hypothesis. Note that we do not exclude possible relationship between these two hypotheses based on the recent pilot study of Gjini et al. (2010) that suggested to use a battery of auditory evoked potential tests to investigate the relationship between electrophysiological measures of salience detection and sensory gating in schizophrenia patients.

The experimental design of this study was based on sound evaluation from participants following three main labels: *bizarre*, *familiar,* and *invasive*. We also included two other dimensions (labeled *frightening* and *reassuring*) related to the emotional features of the stimuli. These labels were chosen to be easily comprehensive by schizophrenic patients. The dimension bizarre corresponded to the feeling of bizarreness, nonsense, unusualness, or distortion induced by sounds. This dimension was investigated in some perceptual scales such as the CAPS (Bell et al., 2006), which included items about inherently unusual or distorted sensory experiences, such as, "Do you ever find that sounds are distorted in strange or unusual ways?" On this CAPS item, the distress, intrusiveness, and frequency of the feeling of bizarre were rated on a Likert scale. The dimension familiar corresponded to the feeling of familiarity, meaning, significance, or usualness induced by sounds. This dimension was explored by Tuscher et al. (2005), who evaluated the familiarity for environmental nonverbal sounds. The dimension invasive corresponded to the "feeling of being flooded/inundated by real sounds" as denoted by a Likert item in the SIAPA (Bunney et al., 1999). Finally, the dimensions reassuring and frightening corresponded to positive and negative feelings induced by sounds. We here preferred these labels rather than the emotional valences commonly used in studies on emotional recognition (i.e., *pleasant* and *unpleasant*). This choice was based on considerations from informal listening pretests, which revealed that the reassuring and frightening labels were easier comprehensive than pleasant and unpleasant during the evaluation of abstract sounds.

We compared sound evaluations on these five perceptual dimensions between groups of schizophrenic patients (SCZ) and healthy subjects (CTL). The perceptual judgment was rated on independent linear scales (one scale for each dimension) for each sound through a computer interface. In addition, the conditioning-testing P50 paradigm was conducted on a subset of schizophrenic patients and control subjects.

Previous studies showed that everyday life sounds were experienced as more strange and unusual (Bell et al., 2006), less familiar (Tuscher et al., 2005), and more flooding/inundating (Bunney et al., 1999) by schizophrenic patients than by healthy subjects due to a sensory gating deficit. Thus, at a behavioral level, we hypothesized that SCZ would evaluate environmental sounds as less familiar and more bizarre. Moreover external meaningless sounds might be experienced as more significant due to an increased attribution of meaning (Cicero et al., 2010). Thus we hypothesized that SCZ would evaluate abstract sounds as more familiar than CTL. In addition, we hypothesized that both types of sounds would be perceived more invasive by SCZ.

At a neurophysiological level, we expected sensory gating deficits in SCZ evaluated by the auditory P50 paradigm based on the previous findings (de Wilde et al., 2007; Patterson et al., 2008). In particular, we expected a positive correlation between invasiveness and S2/S1 amplitude ratio (Bunney et al., 1999) and negative correlations between S1 amplitude (related to the gating in deficit) and familiarity and also between S1 amplitude and bizarreness (Brenner et al., 2009).

Finally, we assumed that hypothesized group differences found in the evaluation of bizarreness, familiarity, invasiveness, or all, might be independent of emotional dimensions. Thus, a lack of differences between groups concerning reassuring and frightening dimensions would confirm results from Tuscher et al. (2005), using a broad range of nonverbal environmental sounds.

Method

Participants

Twenty chronic in and outpatients with schizophrenia from the Department of Psychiatry of Marseille University Hospital, France, constituted the SCZ group. *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.; *DSM–IV*; American Psychiatric Association, 1994) criteria based on Structured Clinical Interview for *DSM–IV* (SCID) interviews assured diagnosis of schizophrenia (First, Gibbon, & Williams, 1997). The patients' clinical severities of illness were assessed by the Scale for the Assessment of Negative Symptoms (SANS) and the Scale for the Assessment of Positive Symptoms (SAPS; Andreasen & Olsen, 1982). Scores were computed from the SANS and the SAPS for a negative symptom factor (mean of affective flattening, avolition/apathy, and anhedonia/asociality), for a psychotic symptom factor (hallucinations and delusions) and for a disorganized symptom factor (positive formal thought disorder and bizarre behavior). The

mean chlorpromazine equivalent dose was calculated according to Davis (1976).

Twenty healthy subjects screened for any current or lifetime history of a *DSM–IV* Axis I disorder based on the Mini-International Neuropsychiatric Interview (MINI; Sheehan et al., 1998) constituted the CTL group. Healthy subjects were matched to patients on the basis of age, gender, handedness (Oldfield, 1971), personal education (years), and the audition habits (hours/days of music listening and playing musical instrument or not). Demographic characteristics for both groups are shown in Table 1.

Exclusion criteria were neurological illness, brain injury, or other significant medical illnesses, current or past substance abuse or dependency, auditory impairment (assessed by a screening audiogram), and prolonged exposure to a language other than French as a child. After complete description of the study to the participants, written informed consent was obtained. The hospital's ethical committee approved the study.

Stimuli: Environmental and Abstract Sounds

We designed a representative corpus of 26 sounds to explore the five perceptual dimensions. Note that the number of sounds was limited to adapt the duration of the experience to schizophrenic patients and to avoid decline of attention and motivation during the test. For that, we started with a larger initial corpus of 199 sounds, including both environmental nonverbal sounds and abstract sounds. Environmental sounds were selected from recorded samples (which sources can be easily recognized, like sounds of animals or waves on the beach) whereas abstract sounds were selected from a sound data bank used by electroacoustic music composers obtained by particular recording techniques or sound transformations. These sounds were equalized by gain adjustments to minimize influence of loudness variations in the sound evaluation.

Then, 7 healthy participants (2 women and 5 men; mean age = 37 years old, $SD = 10.80$), who did not belong to the CTL, evaluated these 199 sounds on the five dimensions by using the same graphical interface designed for the subsequent formal experiment (see description in the Procedure and User Interface section). Results showed a significant negative correlation between familiar and bizarre ($r = -0.70$, $p < .05$), and between reassuring and frightening ($r = -0.56$, $p < .05$) ratings. On the basis of these correlations, we assumed that sounds could be suitably represented in a reduced orthogonal 3D space, which axes corresponded to the bipolar dimensions familiar/bizarre (*x*-axis), reassuring/frightening (*y*-axis), and to the unipolar dimension, invasive (*z*-axis). Sound coordinates along the *x*-axis (*y*-axis, respectively) corresponded to the average ratings of familiar and bizarre (reassuring and frightening, respectively). Sound coordinates along the *z*-axis corresponded to the average ratings for invasive. The 26 final sounds were selected to sample at best this 3D space. For that, we considered the virtual parallelepiped formed by the space occupied by these sounds, and we defined some target positions located on its vertices and on the median positions along its edges and faces. The sounds that were closest to these target positions (in terms of Euclidian distance) were selected. This method ensured the achievement of an optimal final corpus representative of the sound space.

The mean sound duration of the final 26 sounds was 3.66 s ($SD = 1.93$). Sounds were processed in digital format for standardized quality and presentation conditions (sampling frequency 44.1 kHz, 16 bit, mono). Sounds are available at http://www.lma.cnrs-mrs.fr/~kronland/Bizarre/sounds.html.

Procedure and User Interface

The experiment was conducted in a quiet room, where participants were seated in front of a PC computer screen. Sounds were randomly presented using the internal sound card of the computer and open headphones (HD650 Sennheiser) amplified with Samsom (s-type amp). Participants were free to adjust the intensity level of the sounds, once at the beginning of the test. The experimenter made sure that the sound level was high enough to provide a comfortable listening condition for each subject and that this

Table 1
Demographic and Audition Habits Data for Schizophrenic (SCZ) and Control (CTL) Groups Presented as Mean Values (±SD of the M), Except When the Population Number is Specified

Variable	SCZ ($n = 20$)	CTL ($n = 20$)	t^2	p value
Demographic				
Age (in years)	31.90 (6.95)	29.65 (10.87)	3.96	.440
Gender (no. of women–men)	6/14	6/14	—	1
Education (years)	6.40 (3.12)	7.80 (2.31)	2.25	.124
Handedness (no. of participants)				
Right	17	18	—	.643
Left	2	1	—	—
Mixed	1	1	—	—
Audition habits				
Music listening (hours by day)	2 hr (1.34)	1 hr 44' (0.98)	1.51	.463
Musicians (no. of participants)	6	2	—	.235
Hearing threshold (dB)				
Right	14.88 (4.42)	13.08 (3.85)	0.507	.347
Left	14.94 (4.43)	12.89 (4.25)	0.227	.231

Note. Means were compared with *t* tests ($df = 38$) and with chi-square tests for categorical variables. Hearing threshold was obtained from screening audiogram.

level was approximately set to the same value across participants in both groups. The experimenter was present with the participant during the entire test. He followed the process through his own headphones connected on the same Samsom amplifier and noted behavior and reports from the participant.

The experiment began with a six-trial training session[1] to familiarize participants with the task and to ensure the comprehension of meaning of each perceptual dimension. Then, the 26 sounds were randomly presented in a single session. Participants were asked to listen to each sound and to evaluate the familiar, bizarre, invasive, reassuring, and frightening aspects of the sound by positioning a slider on a continuous linear scale (represented by a vertical bar) associated with each perceptual dimension. Each response scale was ranged between two numeric anchors located at the extremities from 0 (*not familiar*) to 100 (*very familiar*) and the label of the perceptual dimension was displayed below the scale. A graphical user interface developed with the Max/MSP software[2] (http://www.cycling74.com/, Max/MSP) was specifically designed for this experiment. Positions of dimension labels displayed on the screen were randomly balanced across participants who could listen to the sound as many times as they wanted by clicking on the *play* button. When participants gave their five ratings of a sound, they switched to the following sound by clicking on the *next* button. Participants could return to previous ratings by clicking on the *previous* button. Although no time constraint was imposed, the experimenter verified that subjects did not dwell too much on each trial. Ratings on the five perceptual dimensions, number of times each sound was listened to, and the duration of the test were collected for each sound and for each subject.

Following the sound evaluation, participants engaged in a brief informal interview with the experimenter where they reported their possible strategies to evaluate sounds [e.g., "Did you have the feeling that you used a specific strategy? Did it happen that you gave a verbal label to sounds? Did you feel like responding before the end of the sounds? Did you feel that the perception of familiarity or bizarreness (popped out from the sounds) was immediate? or did you have to think about it?"]. During these interviews, participants were given the possibility to listen to sounds assessed during the sound evaluation.

ERP Recording

Half of the population of each group (i.e., 10 SCZ and 10 CTL) received auditory ERP recordings. The population was restricted because of the availability of the apparatus in the neurophysiology department. According to the clinical schedule, the ERP recording and the sound evaluation experiment was performed the same week. Subjects were asked to abstain from cigarette smoking for 1 hr before the electrophysiological measurements.

Auditory stimuli were delivered in a conditioning-testing P50 paradigm, consisting of a click pair presentation (conditioning click S1 followed by testing click S2) in a passive task. The interstimulus interval was set to 500 ms, and the interpair interval was set to 10 s. Clicks were rectangular pulses of 50 μs and of intensity of 100 dB SPL (Baker et al., 1987). A set of 60 stimulus pairs was delivered. Subjects were instructed to relax with their eyes closed.

Electroencephalographic activity (EEG) was monitored on a computer (EB Neuro, Inc.). EEG measurements were recorded from one electrode located on the vertex (Cz) and from two electrodes placed on left and right earlobes at 1,000-Hz sampling frequency. The EEG was referenced to the average of right and left earlobes and filtered with a bandpass filter of 1–200 Hz. Data were segmented in single trials of 1,200 ms, starting from 200 ms before the S1 onset and were averaged all together. Electrooculographic data were recorded, and trials contaminated by ocular movements and movement artifacts were rejected by visual inspection.

Statistical Analysis

Data were analyzed using STATISTICA software (Version 7.1, StatSoft). To better examine our hypotheses, we defined subsets of sounds from the initial sound corpus according to the familiar (usual) and bizarre (unusual) dimensions before conducting the statistical analyses. For that, ratings from the CTL group were averaged across participants for these dimensions and were transformed in a 26×26 dissimilarity matrix computed as a sound distance matrix (euclidean norm). Then, a hierarchical cluster analysis using the ward method (Schielke, Fishman, Osatuke, & Stiles, 2009) was conducted. The obtained dendrogram allowed distinguishing three clusters corresponding to subsets of sounds. The first cluster was named *environmental sounds* because these sounds were judged as most familiar and were exclusively environmental sound recordings. The second cluster was named *abstract sounds* because these sounds were judged as most bizarre and were exclusively abstract sounds. In between, the last cluster was named *intermediate sounds* because these sounds were judged no more familiar than bizarre and constituted both environmental and abstract sounds.

Then, ratings were averaged across sounds for each perceptual dimension and for each sound subset. A repeated-measures analysis of variance (ANOVA) was conducted on these averaged ratings including dimension (familiar, bizarre, invasive, reassuring, and frightening) and sound category (environmental sounds, intermediate sounds, and abstract sounds) as within-subject factor, and group (SCZ and CTL) as between-subjects factor. The sources of significant interactions between factors were further examined by *t* tests conducted for each within-subject factor separately. A repeated measures ANOVA was also conducted on the average number of times to which each sound was listened, including sound category as within-subject factor and group as between-subjects factor.

Electrophysiological data were square roots transformed to approximate the normal distributional assumptions required by para-

[1] Six supplementary sounds were chosen in addition to the initial corpus of 199 sounds: two sounds evaluated as very familiar (described as "a bleating of a sheep" and "a Ping-Pong sound"), two sounds evaluated as not very bizarre (described as "like an electronic spring" and "like a transformative fly"), and two sounds evaluated as very bizarre (described as "a drone-like musical sound" and "a very noisy sound").

[2] The interface was tested on 3 healthy subjects and 3 schizophrenic individuals, who did not belong to the SCZ and CTL groups. We controlled that (a) they well understood the task and in particular the meaning of the five perceptual dimensions; (b) they easily answered with the interface; and (c) data were recorded, correctly stored, and easy to export toward a statistic software.

metric statistical methods. Then, data were compared between groups by *t* tests.

Spearman's rank-order correlation was used to examine the relationships between ratings and clinical data and between ratings and electrophysiological data in SCZ. The analyses were considered separately for each sound category, and the correlations were conducted only for ratings that were significantly different between groups (based on the results from the repeated-measure ANOVA). Clinical data were negative symptom, psychotic symptom, and disorganized symptom factors and were correlated with ratings of familiar, bizarre, and invasive. According to our expectations at electrophysiological level, correlations were analyzed between S1/S2 amplitude ratio and rating of invasive. Correlations were also analyzed between S1 amplitude and ratings of familiar, bizarre and invasive. Bonferroni correction was applied if more than one correlation test was performed.

For all analyses, effects were considered significant if the *p* value was equal to or less than .05 (*p* values were reported after the Greenhouse–Geisser correction for nonsphericity).

Results

Demographic and Clinical Data

CTL group did not significantly differ from SCZ group for the demographic and the audition habits data (*t* statistics in Table 1). SCZ patients were neither disorganized nor catatonic, but the following *DSM–IV* subtypes were observed: 12 were paranoid, 4 were undifferentiated, and 4 were residual. All patients were medicated, and the neuroleptic dose was 368.58 mg per day (SD = 218.54; chlorpromazine equivalents; Davis, 1976). Five patients were on typical neuroleptics, 13 were on atypical neuroleptics, 1 was on typical and atypical neuroleptics, and 1 was nontreated. Four patients were on anticholoinergic agents. The mean age for the first psychiatric hospitalization was 20.80 years (SD = 2.67), the mean duration of illness was 11.05 years (SD = 6.75), the total number of hospitalizations was 7.05 (SD = 6.78), and the mean total hospitalization duration was 21.92 months (SD = 42.82). The mean score to the SANS was 48.10 (SD = 24.67), the mean score to the SAPS was 42.95 (SD = 30.46), the mean negative symptom factor was 8.57 (SD = 4.62), the mean psychotic symptom factor was 10.80 (SD = 7.93), and the mean disorganized symptom factor was 7.2 (SD = 5.47).

Behavioral Data: Perceptual Experiences

No significant difference in test duration was found between SCZ (6.85 min, SD = 3.01) and CTL (8.96 min, SD = 6.28; $t^2(38) = 2.42$, $p = .13$). Analysis either revealed no significant difference in the number of times each sound was listened to for each Sound category (Sound Category × Group interaction, $F(2, 76) = 0.55$, $p = .65$. By contrast, ANOVA conducted on averaged ratings revealed a significant effect of the Dimension × Sound Category × Group interaction, $F(8, 304) = 6.65$, $p < .001$. Because the interaction was significant, we reported results from separated analyses for each dimension and for each sound category (*t* statistics in Table 2).

Environmental sounds were evaluated significantly less familiar and more bizarre by SCZ than by CTL (note the exactly null rating of bizarre by all CTL subjects). Intermediate sounds were evaluated no more familiar than bizarre in both groups. Abstract sounds were evaluated significantly more familiar by SCZ than by CTL and similarly bizarre in both groups. Finally, only the category of environmental sounds was judged more invasive by SCZ than by CTL (see Figure 1). Concerning the emotional dimensions, *environmental sounds* were evaluated equally reassuring in both groups and tended to be evaluated as more frightening by SCZ

Table 2
Average Ratings Presented as Mean Values (±SD of the M) and Differences (Schizophrenia–Control [SCZ-CTL]) for Each Sound Category (Environmental Sounds, Intermediate Sounds, and Abstract Sounds), for Each Dimension and for SCZ and CTL Groups

Dimension	SCZ (n = 20)	CTL (n = 20)	Difference (SCZ–CTL)	t^2	p value
Environmental sounds					
Familiar	43.56 (30.24)	66.38 (24.96)	−22.82	6.77	.013*
Bizarre	6.77 (11.41)	0 (0)	6,77	7.05	.011*
Invasive	16.25 (22.21)	5.38 (8.01)	10.87	4.24	.046*
Reassuring	29.92 (22.03)	21.92 (20.44)	8	1.41	.241
Frightening	19.15 (16.83)	11.01 (8.07)	8.14	3.80	.058
Intermediate sounds					
Familiar	26.07 (24.62)	16.84 (16.39)	9.23	1.94	.170
Bizarre	22.84 (17.54)	17.65 (15.92)	5.19	0.95	.333
Invasive	22.48 (17.45)	16.70 (12.63)	5.78	1.43	.237
Reassuring	10.99 (12.38)	2.90 (6.69)	8.09	6.59	.014*
Frightening	22.01 (15.91)	21.65 (10.56)	0.36	.0062	.937
Abstract sounds					
Familiar	12.91 (12.92)	4.53 (8.14)	8.38	6.02	.018*
Bizarre	36.73 (18.38)	39.80 (21.15)	−3.07	0.24	.627
Invasive	31.52 (19.27)	35.72 (21.34)	−4.2	0.44	.518
Reassuring	8.24 (12.40)	1.13 (3.13)	7.11	6.18	.017*
Frightening	27.24 (22.28)	25.10 (17.20)	2.14	0.12	.735

Note. Means were compared with *t* tests (df = 38).
* $p < .05$.

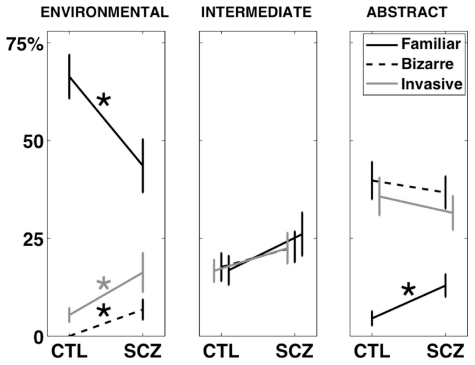


Figure 1. Comparison of averaged ratings on *familiar* (solid line), *bizarre* (dotted line), and *invasive* (gray line) dimensions for the three sound categories ("environmental sounds," "intermediate sounds," and "abstract sounds") between schizophrenia (SCZ; $N = 20$) and control (CTL; $N = 20$) groups. Error bars represent the standard error. * $p < .05$.

than by CTL. Intermediate and abstract sounds were judged significantly more reassuring by SCZ than by CTL and similarly frightening in both groups.

Electrophysiological Data: Sensory Gating Inhibition

The conditioning P50 component (elicited by stimulus S1) was identified as the most positive peak between 40 and 80 ms after the S1 onset (with amplitude >0.5 μV) (Cardenas, Gerson, & Fein, 1993). The P50 component (elicited by stimulus S2) was identified in a similar way after the S2 onset. The amplitudes of these components, called the S1 and S2 amplitudes, respectively, were defined as the difference between the peak of the P50 and the immediately preceding negative peak (Boutros & Belger, 1999). No significant difference was found on the S1 amplitude (SCZ = 2.17 μV, $SD = 1.29$; CTL = 2.52 μV, $SD = 1.66$), $t^2(18) = 0.60$, $p = .604$, and the S2 amplitude (SCZ = 0.91 μV, $SD = 0.91$; CTL = 0.41 μV, $SD = 0.28$), $t^2(18) = 2.80$, $p = .111$, between SCZ and CTL, despite a trend in the S2 amplitude to be larger in SCZ than in CTL. There was no significant difference on the latency of the conditioning P50 (S1 latency, SCZ = 60.43 ms, $SD = 3.11$; CTL = 58.09 ms, $SD = 3.17$), $t^2(18) = 2.75$, $p = .114$, and of the testing P50 (S2 latency, SCZ = 61.67 ms, $SD = 6.44$; CTL = 59.64 ms, $SD = 7.13$), $t^2(18) = 0.46$, $p = .513$, between SCZ and CTL, despite a trend in the S1 latency to be longer in SCZ than in CTL.

In addition, a testing/conditioning ratio (noted S2/S1 ratio) was defined as the ratio between S1 and S2 amplitudes. Low ratios are assumed to reflect an inhibition of sensory gating whereas high ratios (superior to 0.5) may indicate a deficit in the sensory gating. The mean S2/S1 ratio was significantly greater in SCZ (0.46, $SD = 0.37$) than in CTL (0.19, $SD = 0.13$), $t^2(18) = 4.50$, $p = .048$, with half of the SCZ patients who presented a S2/S1 ratio greater than 0.5. By contrast, none of the CTL subject presented a S2/S1 ratio greater than 0.5.

Correlation Analysis

As previously mentioned in the Statistical Analysis section, the correlations were conducted only for ratings that were significantly different between groups (see Table 2) in each sound category. Thus, concerning electrophysiological data, the correlation between S2/S1 ratio and invasive rating was examined for environmental sounds. In addition, the correlation between S1 amplitude and familiar rating was examined for environmental and abstract sounds. Finally, the correlation between S1 amplitude and bizarre and invasive were examined for environmental sounds (significant level at .012 after Bonferroni correction for correlation tests, including S1 amplitude).

First, S2/S1 ratio value positively correlated with the invasive rating for environmental sounds ($r = .695$, $n = 10$, $p = .026$): the more the patients present a deficiency in sensory gating measured with the conditioning-testing P50 paradigm, the more they judge environmental sounds as invasive (Figure 2A). Second, the S1 amplitude negatively correlated with the *familiar* rating for abstract sounds ($r = -.770$, $n = 10$, $p = .009$): the smaller the S1 amplitude, the more patients evaluated abstract sounds as familiar (Figure 2B). No significant correlations with S1 amplitude was found for environmental sounds.

Concerning clinical data, the correlations with familiar rating were conducted for environmental and abstract sounds. The correlations with bizarre and invasive ratings were examined for environmental sounds (significant level at .012 after Bonferroni correction). We found that the intensity of psychotic symptom factor negatively correlated with the evaluation of familiar for environmental sounds in SCZ ($r = -.595$, $n = 20$, $p = .006$). Supplementary analysis revealed that delusion was the most correlated symptom ($r = -.591$, $n = 20$, $p = .007$): the more the patients have delusions the less they evaluated environmental sounds as familiar. Clinical factors did not significant correlate with *familiar* for "Abstract sounds", and *bizarre* and *invasive* dimension for "Environmental sounds".

Discussion

The use of complex sounds allowed us to explore the "feeling of strangeness" previously described in Bell et al. (2006) and more generally abnormal perceptual experiences in SCZ. The selection of stimuli and labels was carefully conducted, and the sound clustering allowed us to accurately investigate the perception of environmental sounds (meaningful), abstract sounds (meaningless), and intermediate sounds. Perception of emotional dimensions (reassuring and frightening) was also examined to ensure that the group differences on the rating of familiar, bizarre, and invasive were not due to any associated emotional effect (Tremeau et al., 2009).

No group difference was found for intermediate sounds for familiarity, bizarreness and invasiveness, and the corresponding results would not be further discussed. By contrast, environmental sounds and abstract sounds were perceived differently by SCZ and CTL. Our findings confirmed some of our expectations according to the two complementary hypotheses presented in the introductory section. The first one (i.e., the sensory gating deficit hypothesis) suggested that abnormal experiences in schizophrenia consist in perceiving things in a distorted, unshared manner, and flooding/inundating, underlying a difficulty in encoding (gating in), or filtering (gating out) sensory inputs (Bell et al., 2006; Bunney et al., 1999). Note that the perceptual deficits observed in SCZ may be related to

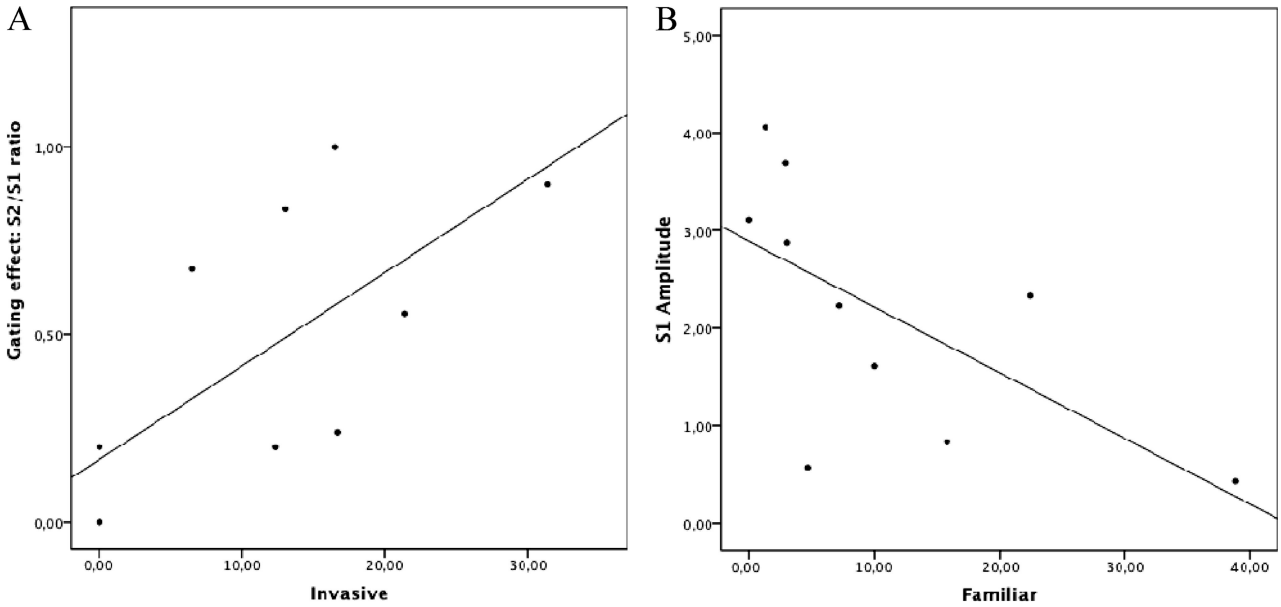


Figure 2. Electrophysiological data in 10 schizophrenic patients. A: Correlation between P50 "gating out" evaluated by the S2/S1 ratio, and evaluation of *invasiveness* on "environmental sounds." A 0 ordinate value corresponds to 100% suppression of the P50 after the second click (S2), and 1 corresponds to absence of diminution of the P50 after the second click (S2). Note that 2 patients have zero values in ordinate and abscissa. $R^2 = .466$. B: Correlation between P50 "gating in" evaluated by the S1 amplitude, and evaluation of *familiar* on "abstract sounds". $R^2 = .401$.

impairment in the processing of timbre features, leading to processing sounds in a distorted manner. This assumption was in line with our previous findings, supporting an impairment in the processing of specific timbre features for material perception (study conducted with impact sounds on metal, wood, and glass materials) in schizophrenia (Micoulaud-Franchi et al., 2011). The second hypothesis (i.e., the aberrant salience hypothesis) suggested that abnormal experiences consist in perceiving things that commonly have no particular meaning as things that become meaningful by assigning them incorrect significances (Cicero et al., 2010; Kapur, 2003). In addition, our finding brought some perspectives concerning ambivalence that may not only be related to these previous hypotheses but also more generally to the assumption that abnormal experiences may also be due to impairment in high levels of sound processing (in particular, semantic identification impairment) and to the coactivation of conflicting processes in schizophrenia.

Abnormal Perceptual Experience of Environmental Sounds

As expected, SCZ evaluated environmental sounds as less familiar, more bizarre, and more invasive than CTL. These results were in agreement with patients' perceptual reports obtained by SIAPA or CAPS, indicating that everyday life sounds seem to be experienced as distorted, unusual, nonshared, and flooding/inundating (Bell et al., 2006; Bunney et al., 1999). In addition, we found that the rating of familiar for environmental sounds was negatively correlated with the intensity of clinical symptoms (psychotic symptom factor and particularly delusion). This result was in line with previous studies, showing that anomalous perceptual experiences might enhance delusion (Bilder, Mukherjee, Rieder, & Pandurangi, 1985; Peralta & Cuesta, 1999; Peralta, de Leon, & Cuesta, 1992). Nevertheless, this aspect remains controversial and some other studies that did not report such a relationship using the CAPS (Bell, Halligan, & Ellis, 2008).

A significantly greater S2/S1 ratio (recorded from conditioning-testing P50 paradigm) was found in SCZ. Moreover, this ratio was positively correlated with the invasive rating for environmental sounds, constituting valuable arguments in favor of the sensory gating deficit hypothesis. More precisely, impairments in the sensory gating out process of irrelevant sensory inputs might cause a perceptual experience of inundation and flooding in this disease (Boutros & Belger, 1999; Brenner et al., 2009; Bunney et al., 1999). Our results contrasted with those of Jin et al. (1998), who did not find any relationship between perceptual anomalies assessed by the SIAPA and sensory gating deficit assessed by P50 recording. As in Light and Braff (2000), this difference might be explained by a deficiency of insight and self-awareness from patients' self-reports investigated by Jin et al. (1998). This study avoids such a possible drawback, because the evaluation of invasiveness was done during sound listening and thus diminishing the confounding effects of altered insight and self-awareness (Light & Braff, 2000), as well as the avoidance of perceived stigmatization (Kruck et al., 2009).

Abnormal Perceptual Experience of Abstract Sounds

As partly expected, we found that SCZ evaluated abstract sounds as more familiar but no more invasive than CTL. The overevaluation of familiarity indicated that patients might experience meaningless sounds as more significant than healthy subjects in agreement with schizophrenic individuals' perceptual reports by using the ASI scale (Cicero et al., 2010) and with the aberrant salience hypothesis developed by Kapur. This result might be accurately compared with previous ones from Nielzen, Olsson, and Ohman (1993), who demonstrated that patients judged complex nonverbal nonenvironmental sounds as more attractive than healthy subjects.

In addition, we found a negative correlation between the familiar rating for abstract sounds and S1 amplitude in SCZ. This negative correlation indicated that aberrant salience anomalies (revealed by a higher rating of familiarity for meaningless sounds) seemed to be enhanced by the deficit to gate in novel sensory inputs (revealed by the S1 amplitude decrease). Previous studies already concluded on a relationship between gating in deficit and phenomenological perceptual anomalies but remained attached to the assumption of distorted, unusual, and unshared perception in line with the sensory gating hypothesis (Hetrick et al., 2010; Johannesen et al., 2008; Kisley, Noecker, & Guinther, 2004). As a complement to these previous studies, our results also suggested a relationship between gating in deficit and emergence of aberrant saliencies in line with the aberrant salience hypothesis. Thus, our results confirmed that a better characterization of this relationship could improve the understanding of the sensory-processing abnormalities in schizophrenia (Gjini, Arfken, & Boutros, 2010).

The abstract sounds were evaluated as highly flooding similarly between CTL and SCZ. These results might be explained because the consequence of sensory gating deficit might be hidden by a saturation effect for very invasive sounds. Note that it was not the case for environmental sounds because the rating of invasive was notably lower.

Ambivalence in Abnormal Perceptual Experience

We designed the experimental protocol so that familiarity and bizarreness were evaluated in a same trial (for each sound) but separately on two distinct scales. We observed that CTL evaluated these dimensions in a quite categorical way: Environmental sounds were perceived as highly familiar and not bizarre at all (exactly null rating) and abstract sounds were perceived as mostly bizarre and almost non familiar (exact null rating for most subjects except for only a few). By contrast, we found that patients were not disturbed by evaluating sounds jointly as familiar and bizarre: Environmental sounds were mainly perceived as familiar (even if it was lower than CTL) and more bizarre than CTL; abstract sounds were mainly perceived as bizarre (similarly to CTL) and more familiar than CTL.

Beyond the two hypotheses examined in our study, we assume that these results were due to a more complex process than the ability to gate in–out relevant sensory inputs or to attribute significance to aberrant auditory saliencies. We suggested that they might be linked to the concept of "ambivalence" in schizophrenia defined as "the tendency. . . to endow the most diverse psychisms with both a positive and negative indicator at one and the same time" (Bleuler, 1950, p. 53). Ambivalence in schizophrenia was previously demonstrated for emotional recognition processing, using calibrated stimuli of the International Affective Digitized Sounds: Patients rated positive stimuli as pleasant and negative stimuli as unpleasant similarly to healthy subjects, but at the same time they rated positive stimuli as more unpleasant, and negative stimuli as more pleasant (Lang, Bradley, & Cuthbert, 1999; Tremeau et al., 2009). This result was formally explained by the coactivation of the two emotional evaluative systems (one for positivity and one for negativity) assumed to be opposite and distinct in healthy subjects (Tremeau et al., 2009). In our study, emotional dimensions were globally well controlled since ratings of reassuring and frightening were similar in both groups for environmental sounds and almost for abstract sounds (excepted the rating of reassuring that was greater in SCZ). These results allowed us to conclude that the evaluation of bizarreness and familiarity (and also invasiveness) was not much influenced by emotional dimensions. Thus, we assumed that our result on joint and nonnull evaluation of familiar and bizarre in SCZ might even so be interpreted as ambivalence effect but not based on two emotional evaluative processing. In particular, we assumed that the observed ambivalence might be related to consequences of an impairment in the semantic process of labeling, indicating a conflicting coactivation of two types of listening,[3] that is, "everyday listening" and "musical listening" as defined by Gaver (Gaver, 1993). In our view, this last term of musical listening is close to the notion of acousmatic listening, which is used in our study.

Actually, the underrating of familiar for environmental sounds in SCZ supported the assumption of impairment in semantic process of labeling in line with previous studies. Contrary to CTL, SCZ reported that, "For all sounds, I asked myself whether I could label this sound?" indicating that the labeling may not be an automatic or easy process for them. Moreover, Tuscher et al. explained the underevaluation of familiarity for environmental sounds by patients compared to healthy subjects as a consequence of an inability to activate the appropriate representation in the internal lexicon memory when a stimulus is perceived (Tuscher et al., 2005). Wexler et al.'s (2002) study mainly focused on semantic processing and showed deficit in schizophrenia in verbal memory processes independently from sensory processing. At least, semantic impairments were found in schizophrenia both with a priming protocol in a lexical-decision task (Spitzer, 1997), and with the N400 protocol (ERP method) considered as a neurophysiological probe of activation of concepts in semantic memory (Kiang, Kutas, Light, & Braff, 2008). Going further, it was suggested that this semantic impairment in schizophrenia might be due to the development of a hyperextended semantic network as proposed by Spitzer (1997). The recourse to this network may be emphasized for abstract sounds, defined to be particularly difficult to label (even for CTL) because they could not be easily associated with a physical sound source.

[3] Note that Schaeffer previously proposed four types of listening, that is, "hearing," "listening," "comprehending," and "understanding." Some of them were in line with the ones proposed by Gaver. The four types of listening involved subtle differentiations between cognitive processes and may not be adapted to directly support our findings.

In relation with this semantic impairment and possible sensory deficient, we suggested that SCZ might coactivate the two types of listening, that is, everyday listening and acousmatic listening, in a conflicting way. Everyday listening refers to a listening of sounds as things. Ihde wrote: "Sounds are 'first' experienced as sounds of things and it is sufficient for ordinary affairs" (Ihde, 1976, p. 60), and these things may consistently be associated with labels. Thus, everyday listening necessitates a labeling process of sounds. Acousmatic listening refers to the listening of sounds in terms of their sound quality (acoustic attributes) without paying attention to the sources. Schaeffer wrote, "Often surprised, often uncertain, we discover that much of what we thought we were hearing, was in reality only seen, and explained, by the context" (Schaeffer, 1966, p. 93). Generally, sound listening leads to switching from one type of listening to another. As detailed in the Introduction and in our previous research on healthy subjects (Petitmengin et al., 2009), "Abstract sounds" may privilege "acousmatic listening" and "Environmental sounds" may privilege "everyday listening".

On the basis of our findings concerning environmental sounds, we assumed that patients tried to label these sounds by connecting them to things using everyday listening but had difficulties in activating the appropriate labels combined with a distorted, nonshared perception that also might have led them to adopt acousmatic listening. In line with this assumption, we observed that patients had some trouble to label sounds, as shown by patients' reports, reported above. In addition, patients also reported more details concerning sound quality, such as volume, pitch, or timbre attributes, such as, "This sound of wind presents a beautiful tremolo." Concerning abstract sounds, our study showed a similar evaluation of bizarre in the two groups: We assumed that patients adopted acousmatic listening as healthy participants also might have done. Nevertheless, a possible call in the hyperextended semantic network might also have led them to activate everyday listening in a significant way, allowing them to attribute a conceivable meaning to sounds though the feeling of bizarreness was present. Some patients' reports on abstract sounds illustrated this dual aspect: "It sounds like a chirping bird, but a strange bird in a science fiction movie"; "It could be the sound of a leaky faucet, but a strange faucet." Moreover, we observed that patients consistently tried to find a plausible description of abstract sounds, by frequently referring to science fiction productions because of a lack of real-life references: "This is like the sound of a vessel in an alien movie." Finally, these results and considerations allowed us to conclude that the observed ambivalence in SCZ, in particular the overrating of bizarre for environmental sounds and the overrating of familiar for abstract sounds might be due to the simultaneous activation of these two types of listening.

Limitations

Some limitations can be attributed to our study. First, concerning the experimental design, we used a continuous linear scale ranging from 0 to 100, leading to a higher variability of response values than a discrete Likert scale. Moreover, the use of continuous linear scales may induce a different understanding of the scale rating between SCZ and CTL, meaning that the observed effects might not be solely due to abnormal perception in SCZ but rather to a systematic group difference in the anchors assumed by the two groups. However, during the training session the experimenter ensured that subjects understood the task and the meaning of the perceptual dimension in a similar way. In addition, results showed that the group difference was not significant for several perceptual dimensions and that the variances of the ratings globally were comparable in the two groups (see Table 2). We also found that SCZ rated environmental sounds as less familiar than CTL and as more familiar than CTL for abstract sounds, indicating an opposite effect along this perceptual dimension. Thus, our findings can be reasonably related to perceptual abnormalities in SCZ, hereby reducing the probability of a systematic group difference between CTL and SCZ.

Second, we did not use emotional valences usually experimented in the literature. Because we did not aim at precisely investigating the emotional recognition in schizophrenia, we did not evaluate specifically pleasant and unpleasant feelings induced by sounds. Nevertheless, we chose labels associated with the emotional dimensions that were well adapted to our sound corpus, particularly to the abstract sounds.

Third, we did not specifically analyze hypersensitivity (e.g., SIAPA-item like "real sound seem more intense or loud") or selective attention to external sounds (e.g., SIAPA-item like "cannot focus attention on one real sound by excluding the other ones"; Bunney et al., 1999). Although the design of our study was not created to explore these aspects, the related acoustic characteristics of sounds were controlled at best during the design of the sound corpus. In particular, the experimenter equalized the sound intensity level, and the coexistence of several auditory streams in a sound was minimized to control the subjects' attention. On the basis of previous studies that showed a correlation between attentional performance and sensory gating deficit assessed by P50 response (Cullum et al., 1993; Erwin, Turetsky, Moberg, Gur, & Gur, 1998; Wan, Friedman, Boutros, & Crawford, 2008), our findings concerning sensory gating deficit may also be induced by attentional deficit. However, the duration of the test and the number of times each sound was listened to, were similar between groups, allowing us to conclude that there was a lack of noticeable attentional or motivational disturbance in SCZ.

Fourth, the task conducted in this study involved multiple cognitive processes and our results cannot determine whether group differences can be related to a low level (perceptual), to a high level (cognitive and semantic) or to interactions between low- and high-level dysfunctions in schizophrenia as discussed in previous literature (Adcock et al., 2009; Bell et al., 2008; Javitt, 2009; Leitman et al., 2010). Moreover, based on Phillips and Silverstein's (2003) study, we cannot exclude an alteration in perceptual and cognitive organization and neural synchrony as a possible interpretation of group differences. Actually, this interpretation could subsequently be validated experimentally if the EEG activity reveals a decreased synchrony for environmental sounds and an increased synchrony for abstract sounds in schizophrenia compared with healthy subjects.

As a last limitation, our study suffered from a small population size. However, individual data did not show any outliers among participants, and our results are in line with the existing literature. To support our results, a larger cohort of patients is necessary, and it will be interesting to investigate patients that present emphasized abnormal experiences such as young schizophrenia patients or high-risk persons for schizophrenia (Cicero et al., 2010; Horan,

Reise, Subotnik, Ventura, & Nuechterlein, 2008; Parnas, Handest, Jansson, & Saebye, 2005; Parnas, Moller, et al., 2005).

Conclusion

The principal aim of this study was to investigate the "feeling of strangeness" suggested to be a general feature of schizophrenic perceptual experience (Bell et al., 2006; Blankenburg & Mishara, 2001; Cermolacce et al., 2010; Stanghellini, 2000). The experimental paradigm proposed in this study may be considered as a complementary approach to perceptual scales (Bell et al., 2006; Bunney et al., 1999; Cicero et al., 2010; Hetrick et al., 2010) with the advantage to be less sensitive to the effects of insight and self-awareness alteration as observed in patients' self reports (Kruck et al., 2009; Light & Braff, 2000). In addition, this was the first study to our knowledge that used abstract sounds to assess abnormal perceptual experience in schizophrenia (Merer et al., 2010; Schaeffer, 1966).

By designing a specific sound corpus and task procedure, our findings allowed us to provide arguments in favor of both sensory gating deficit hypothesis and aberrant salience hypothesis. Actually, we can conclude that the abnormal perceptual experience related to feeling of strangeness in schizophrenia is based on two processes: perceiving usual or meaningful things (i.e., environmental sounds) in an unusual or meaningless way (due to a sensory deficit) and perceiving unusual or meaningless things (i.e., abstract sounds) in a meaningful way though the feeling of bizarreness was not avoided. Results also highlighted ambivalence on familiarity and bizarreness in SCZ, with a similar evaluation of bizarre for abstract sounds between SCZ and CTL. Thus, beyond the previous hypotheses, the observed ambivalence could also be compatible with the explanation of semantic process impairment related to a hyperextended semantic network in schizophrenia. We suggested that this ambivalence was due to the coactivation of two types of listening, "everyday" and "acousmatic" listening, in a conflicting way (Gaver, 1993; Schaeffer, 1966).

In conclusion, the use of specific environmental and abstract sounds allowed us to explore perception of complex auditory stimuli in schizophrenia. This approach needs more investigations on patients' reports together with the use of electrophysiological measurement. In particular, ERPs studies offer a precise discrimination between semantic processes and sensory or perceptual processes (Aramaki, Marie, Kronland-Martinet, Ystad, & Besson, 2011; Kiang et al., 2008; Micoulaud-Franchi et al., 2011; Schon et al., 2011). Another promising domain of research consists in the exploration of the synchrony EEG activity, with a better spatial and temporal understanding of the neural discharges involved in a coherent object representation (Lachaux, Rodriguez, Martinerie, & Varela, 1999; Tallon-Baudry & Bertrand, 1999). This study and these perspectives constitute new experimental steps to tackle actual difficulties of patients with schizophrenia to experience everyday situations as familiar, veridical, and shareable.

References

Adcock, R. A., Dale, C., Fisher, M., Aldebot, S., Genevsky, A., Simpson, G. V., . . . Vinogradov, S. (2009). When top-down meets bottom-up: Auditory training enhances verbal memory in schizophrenia. *Schizophrenia Bulletin, 35,* 1132–1141. doi:10.1093/schbul/sbp068

Adler, L. E., Pachtman, E., Franks, R. D., Pecevich, M., Waldo, M. C., & Freedman, R. (1982). Neurophysiological evidence for a defect in neuronal mechanisms involved in sensory gating in schizophrenia. *Biological Psychiatry, 17,* 639–654.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

Andreasen, N. C., Arndt, S., Swayze, V., 2nd, Cizadlo, T., Flaum, M., O'Leary, D., . . . Yuh, W. T. (1994, October 14). Thalamic abnormalities in schizophrenia visualized through magnetic resonance image averaging. *Science, 266,* 294–298. doi:10.1126/science.7939669

Andreasen, N. C., & Olsen, S. (1982). Negative v positive schizophrenia. Definition and validation. *Archives of General Psychiatry, 39,* 789–794.

Aramaki, M., Marie, C., Kronland-Martinet, R., Ystad, S., & Besson, M. (2011). Sound categorization and conceptual priming for nonlinguistic and linguistic sounds. *Journal of Cognitive Neuroscience, 22,* 2555–2569. doi:10.1162/jocn.2009.21398

Baker, N., Adler, L. E., Franks, R. D., Waldo, M., Berry, S., Nagamoto, H., . . . Freedman, R. (1987). Neurophysiological assessment of sensory gating in psychiatric inpatients: Comparison between schizophrenia and other diagnoses. *Biological Psychiatry, 22,* 603–617. doi:10.1016/0006-3223(87)90188-0

Bell, V., Halligan, P. W., & Ellis, H. D. (2006). The Cardiff Anomalous Perceptions Scale (CAPS): A new validated measure of anomalous perceptual experience. *Schizophrenia Bulletin, 32,* 366–377. doi:10.1093/schbul/sbj014

Bell, V., Halligan, P. W., & Ellis, H. D. (2008). Are anomalous perceptual experiences necessary for delusions? *Journal of Nervous and Mental Disease, 196,* 3–8. doi:10.1097/NMD.0b013e31815f6619

Bilder, R. M., Mukherjee, S., Rieder, R. O., & Pandurangi, A. K. (1985). Symptomatic and neuropsychological components of defect states. *Schizophrenia Bulletin, 11,* 409–419.

Blankenburg, W., & Mishara, A. L. (2001). First steps toward a psychopathology of "common sense." *Philosophy, Psychiatry, & Psychology, 8,* 303–315. doi:10.1353/ppp.2002.0014

Bleuler, E. (1950). *Dementia Praecox or the Group of Schizophrenias.* (p. 53). New York, NY: International Universities Press.

Boutros, N. N., & Belger, A. (1999). Midlatency evoked potentials attenuation and augmentation reflect different aspects of sensory gating. *Biological Psychiatry, 45,* 917–922. doi:10.1016/S0006-3223(98)00253-4

Braff, D. L., & Geyer, M. A. (1990). Sensorimotor gating and schizophrenia. Human and animal model studies. *Archives of General Psychiatry, 47,* 181–188.

Brenner, C. A., Kieffaber, P. D., Clementz, B. A., Johannesen, J. K., Shekhar, A., O'Donnell, B. F., & Hetrick, W. P. (2009). Event-related potential abnormalities in schizophrenia: A failure to "gate in" salient information? *Schizophrenia Research, 113,* 332–338. doi:10.1016/j.schres.2009.06.012

Bunney, W. E., Jr., Hetrick, W. P., Bunney, B. G., Patterson, J. V., Jin, Y., Potkin, S. G., & Sandman, C. A. (1999). Structured Interview for Assessing Perceptual Anomalies (SIAPA). *Schizophrenia Bulletin, 25,* 577–592.

Cardenas, V. A., Gerson, J., & Fein, G. (1993). The reliability of P50 suppression as measured by the conditioning/testing ratio is vastly improved by dipole modeling. *Biological Psychiatry, 33,* 335–344. doi:10.1016/0006-3223(93)90322-5

Cermolacce, M., Sass, L., & Parnas, J. (2010). What is bizarre in bizarre delusions? A critical review. *Schizophrenia Bulletin, 36,* 667–679. doi:10.1093/schbul/sbq001

Cicero, D. C., Kerns, J. G., & McCarthy, D. M. (2010). The Aberrant Salience Inventory: A new measure of psychosis proneness. *Psychological Assessment, 22,* 688–701. doi:10.1037/a0019913

Cullum, C. M., Harris, J. G., Waldo, M. C., Smernoff, E., Madison, A., Nagamoto, H. T., . . . Freedman, R. (1993). Neurophysiological and

neuropsychological evidence for attentional dysfunction in schizophrenia. *Schizophrenia Research, 10,* 131–141. doi:10.1016/0920-9964(93)90048-N

Cutting, J., & Dunne, F. (1989). Subjective experience of schizophrenia. *Schizophrenia Bulletin, 15,* 217–231.

Davis, J. M. (1976). Comparative doses and costs of antipsychotic medication. *Archives of General Psychiatry, 33,* 858–861.

de Wilde, O. M., Bour, L. J., Dingemans, P. M., Koelman, J. H., & Linszen, D. H. (2007). A meta-analysis of P50 studies in patients with schizophrenia and relatives: Differences in methodology between research groups. *Schizophrenia Research, 97,* 137–151. doi:10.1016/j.schres.2007.04.028

Erwin, R. J., Turetsky, B. I., Moberg, P., Gur, R. C., & Gur, R. E. (1998). P50 abnormalities in schizophrenia: Relationship to clinical and neuropsychological indices of attention. *Schizophrenia Research, 33,* 157–167. doi:10.1016/S0920-9964(98)00075-9

First, R., Gibbon, M., & Williams, J. (1997). *Structured Clinical Interview for DSM–IV Axis I Disorders: Clinical Version (SCID-CV).* Washington, DC: American Psychiatric Press.

Freedman, R., Adler, L. E., Gerhardt, G. A., Waldo, M., Baker, N., Rose, G. M., . . . Franks, R. (1987). Neurobiological studies of sensory gating in schizophrenia. *Schizophrenia Bulletin, 13,* 669–678.

Freedman, R., Adler, L. E., Waldo, M. C., Pachtman, E., & Franks, R. D. (1983). Neurophysiological evidence for a defect in inhibitory pathways in schizophrenia: Comparison of medicated and drug-free patients. *Biological Psychiatry, 18,* 537–551.

Gaver, W. W. (1993). How do we hear in the world?: Exploration in ecological acoustics. *Ecological psychology, 5,* 28. doi:10.1207/s15326969eco0504_2

Gjini, K., Arfken, C., & Boutros, N. N. (2010). Relationship between sensory "gating out" and sensory "gating in" of auditory evoked potentials in schizophrenia: A pilot study. *Schizophrenia Research, 121,* 139–145. doi:10.1016/j.schres.2010.04.020

Heinz, A., & Schlagenhauf, F. (2010). Dopaminergic dysfunction in schizophrenia: Salience attribution revisited. *Schizophrenia Bulletin, 36,* 472–485. doi:10.1093/schbul/sbq031

Hetrick, W. P., Erickson, M. A., & Smith, D. A. (2010). Phenomenological dimensions of sensory gating. *Schizophrenia Bulletin.* Advance online publication. doi:10.1093/schbul/sbq054

Horan, W. P., Reise, S. P., Subotnik, K. L., Ventura, J., & Nuechterlein, K. H. (2008). The validity of Psychosis Proneness Scales as vulnerability indicators in recent-onset schizophrenia patients. *Schizophrenia Research, 100,* 224–236. doi:10.1016/j.schres.2007.12.469

Ihde, D. (1976). *Listening and voice. Phenomenology of sound.* Athens, OH: Ohio University Press.

Javitt, D. C. (2009). Sensory processing in schizophrenia: Neither simple nor intact. *Schizophrenia Bulletin, 35,* 1059–1064. doi:10.1093/schbul/sbp110

Jin, Y., Bunney, W. E., Jr., Sandman, C. A., Patterson, J. V., Fleming, K., Moenter, J. R., . . . Potkin, S. G. (1998). Is P50 suppression a measure of sensory gating in schizophrenia? *Biological Psychiatry, 43,* 873–878. doi:10.1016/S0006-3223(98)00115-2

Johannesen, J. K., Bodkins, M., O'Donnell, B. F., Shekhar, A., & Hetrick, W. P. (2008). Perceptual anomalies in schizophrenia co-occur with selective impairments in the gamma frequency component of midlatency auditory ERPs. *Journal of Abnormal Psychology, 117,* 106–118. doi:10.1037/0021-843X.117.1.106

Kapur, S. (2003). Psychosis as a state of aberrant salience: A framework linking biology, phenomenology, and pharmacology in schizophrenia. *American Journal of Psychiatry, 160,* 13–23. doi:10.1176/appi.ajp.160.1.13

Kiang, M., Kutas, M., Light, G. A., & Braff, D. L. (2008). An event-related brain potential study of direct and indirect semantic priming in schizophrenia. *American Journal of Psychiatry, 165,* 74–81. doi:10.1176/appi.ajp.2007.07050763

Kisley, M. A., Noecker, T. L., & Guinther, P. M. (2004). Comparison of sensory gating to mismatch negativity and self-reported perceptual phenomena in healthy adults. *Psychophysiology, 41,* 604–612. doi:10.1111/j.1469-8986.2004.00191.x

Kruck, C. L., Flashman, L. A., Roth, R. M., Koven, N. S., McAllister, T. W., & Saykin, A. J. (2009). Lack of relationship between psychological denial and unawareness of illness in schizophrenia-spectrum disorders. *Psychiatry Research, 169,* 33–38. doi:10.1016/j.psychres.2008.07.010

Lachaux, J. P., Rodriguez, E., Martinerie, J., & Varela, F. J. (1999). Measuring phase synchrony in brain signals. *Human Brain Mapping, 8,* 194–208. doi:10.1002/(SICI)1097-0193(1999)8:4<194::AID-HBM4>3.0.CO;2-C

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1999). International affective picture system (IAPS): Technical manual and affective ratings. Gainesville, FL: Center for Research in Psychophysiology, University of Florida.

Leitman, D. I., Sehatpour, P., Higgins, B. A., Foxe, J. J., Silipo, G., & Javitt, D. C. (2010). Sensory deficits and distributed hierarchical dysfunction in schizophrenia. *American Journal of Psychiatry, 167,* 818–827. doi:10.1176/appi.ajp.2010.09030338

Light, G. A., & Braff, D. L. (2000). Do self-reports of perceptual anomalies reflect gating deficits in schizophrenia patients? *Biological Psychiatry, 47,* 463–467. doi:10.1016/S0006-3223(99)00280-2

McGhie, A., & Chapman, J. (1961). Disorders of attention and perception in early schizophrenia. *British Journal of Medical Psychology, 34,* 103–116. doi:10.1111/j.2044-8341.1961.tb00936.x

Merer, A., Ystad, S., Kronland-Martinet, R., & Aramaki, M. (2010, June). *On the potentiality of abstract sounds in perception research.* Paper presented at the 7th International Symposium on Computer Music Modeling and Retrieval, Málaga, Spain.

Micoulaud-Franchi, J. A., Aramaki, M., Merer, A., Cermolacce, M., Ystad, S., Kronland-Martinet, R., & Vion-Dury, J. (2011). Categorization and timbre perception of environmental sounds in schizophrenia. *Psychiatry Research, 189,* 148–152.

Nielzen, S., Olsson, O., & Ohman, R. (1993). On perception of complex sound in schizophrenia and mania. *Psychopathology, 26,* 13–23. doi:10.1159/000284795

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9,* 97–113. doi:10.1016/0028-3932(71)90067-4

Parnas, J., Handest, P., Jansson, L., & Saebye, D. (2005). Anomalous subjective experience among first-admitted schizophrenia spectrum patients: Empirical investigation. *Psychopathology, 38,* 259–267. doi:10.1159/000088442

Parnas, J., Moller, P., Kircher, T., Thalbitzer, J., Jansson, L., Handest, P., & Zahavi, D. (2005). EASE: Examination of Anomalous Self-Experience. *Psychopathology, 38,* 236–258. doi:10.1159/000088441

Patterson, J. V., Hetrick, W. P., Boutros, N. N., Jin, Y., Sandman, C., Stern, H., . . . Bunney, W. E., Jr. (2008). P50 sensory gating ratios in schizophrenics and controls: A review and data analysis. *Psychiatry Research, 158,* 226–247. doi:10.1016/j.psychres.2007.02.009

Peralta, V., & Cuesta, M. J. (1999). Dimensional structure of psychotic symptoms: An item-level analysis of SAPS and SANS symptoms in psychotic disorders. *Schizophrenia Research, 38,* 13–26. doi:10.1016/S0920-9964(99)00003-1

Peralta, V., de Leon, J., & Cuesta, M. J. (1992). Are there more than two syndromes in schizophrenia? A critique of the positive-negative dichotomy. *British Journal of Psychiatry, 161,* 335–343. doi:10.1192/bjp.161.3.335

Petitmengin, C., Bitbol, M., Nissou, J. M., Pachoud, B., Curalucci, H.,

Cermolacce, M., & Vion-Dury, J. (2009). Listening from within. *Journal of Consciousness Studies, 16,* 252–284.

Phillips, W. A., & Silverstein, S. M. (2003). Convergence of biological and psychological perspectives on cognitive coordination in schizophrenia. *Behavioral and Brain Science, 26,* 65–82. doi:10.1017/S0140525X03000025

Sass, L. (2001). Self and world in schizophrenia: Three classic approaches. *Philosophy, Psychiatry, & Psychology, 8,* 251–270. doi:10.1353/ppp.2002.0026

Schaeffer, P. (1966). *Traité des objets musicaux.* Paris, France: Seuil.

Schielke, H. J., Fishman, J. L., Osatuke, K., & Stiles, W. B. (2009). Creative consensus on interpretations of qualitative data: The Ward method. *Psychotherapy Research, 19,* 558–565. doi:10.1080/10503300802621180

Schon, D., Ystad, S., Kronland-Martinet, R., & Besson, M. (2011). The evocative power of sounds: Conceptual priming between words and nonverbal sounds. *Journal of Cognitive Neuroscience, 22,* 1026–1035. doi:10.1162/jocn.2009.21302

Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., . . . Dunbar, G. C. (1998). The Mini-International Neuropsychiatric Interview (M.I.N.I.): The development and validation of a structured diagnostic psychiatric interview for *DSM–IV* and ICD-10. *Journal of Clinical Psychiatry, 59*(Suppl 20), 22–33.

Solomon, L. N. (1958). Semantic approach to the perception of complex sounds. *The Journal of Acoustical Society of America, 30,* 421–425. doi:10.1121/1.1909632

Spitzer, M. (1997). A cognitive neuroscience view of schizophrenic thought disorder. *Schizophrenia Bulletin, 23,* 29–50.

Stanghellini, G. (2000). Vulnerability to schizophrenia and lack of common sense. *Schizophrenia Bulletin, 26,* 775–787.

Tallon-Baudry, C., & Bertrand, O. (1999). Oscillatory gamma activity in humans and its role in object representation. *Trends in Cognitive Sciences, 3,* 151–162. doi:10.1016/S1364-6613(99)01299-1

Tremeau, F., Antonius, D., Cacioppo, J. T., Ziwich, R., Jalbrzikowski, M., Saccente, E., . . . Javitt, D. (2009). In support of Bleuler: Objective evidence for increased affective ambivalence in schizophrenia based upon evocative testing. *Schizophrenia Research, 107,* 223–231. doi:10.1016/j.schres.2008.09.020

Tuscher, O., Silbersweig, D., Pan, H., Smith, T., Beutel, M., Zonana, J., . . . Engelien, A. (2005). Processing of environmental sounds in schizophrenic patients: Disordered recognition and lack of semantic specificity. *Schizophrenia Research, 73,* 291–295. doi:10.1016/j.schres.2004.06.010

Uhlhaas, P. J., & Mishara, A. L. (2007). Perceptual anomalies in schizophrenia: Integrating phenomenology and cognitive neuroscience. *Schizophrenia Bulletin, 33,* 142–156. doi:10.1093/schbul/sbl047

Wan, L., Friedman, B. H., Boutros, N. N., & Crawford, H. J. (2008). P50 sensory gating and attentional performance. *International Journal of Psychophysiology, 67,* 91–100. doi:10.1016/j.ijpsycho.2007.10.008

Wexler, B. E., Donegan, N., Stevens, A. A., & Jacob, S. A. (2002). Deficits in language-mediated mental operations in patients with schizophrenia. *Schzophrenia Research, 53,* 171–179.

Wiggins, O. P., Schwartz, M. A., & Naudin, J. (2001). Husserlian comments on Blankenburg's "Psychopathology of common sense." *Philosophy, Psychiatry, & Psychology, 8,* 327–329. doi:10.1353/ppp.2002.0030

Received June 12, 2011
Revision received September 28, 2011
Accepted October 11, 2011 ■

Prospective View on Sound Synthesis BCI Control in Light of Two Paradigms of Cognitive Neuroscience

4

Mitsuko Aramaki, Richard Kronland-Martinet, Sølvi Ystad, Jean-Arthur Micoulaud-Franchi and Jean Vion-Dury

Abstract

Different trends and perspectives on sound synthesis control issues within a cognitive neuroscience framework are addressed in this article. Two approaches for sound synthesis based on the modelling of physical sources and on the modelling of perceptual effects involving the identification of invariant sound morphologies (linked to sound semiotics) are exposed. Depending on the chosen approach, we assume that the resulting synthesis models can fall under either one of the theoretical frameworks inspired by the representational-computational or enactive paradigms. In particular, a change of viewpoint on the epistemological position of the end-user from a third to a first person inherently involves different conceptualizations of the interaction between the listener and the sounding object. This differentiation also influences the design of the control strategy

M. Aramaki (✉) · R. Kronland-Martinet · S. Ystad
Laboratoire de Mécanique et d'Acoustique (LMA), CNRS UPR 7051, Aix-Marseille University, Centrale Marseille, 31, Chemin Joseph Aiguier, 13402 Marseille Cedex 20, France
e-mail: aramaki@lma.cnrs-mrs.fr

R. Kronland-Martinet
e-mail: kronland@lma.cnrs-mrs.fr

S. Ystad
e-mail: ystad@lma.cnrs-mrs.fr

J.-A. Micoulaud-Franchi · J. Vion-Dury
Laboratoire de Neurosciences Cognitives (LNC), CNRS UMR 7291, Aix-Marseille University, Site St Charles, 3, Place Victor Hugo, 13331 Marseille Cedex 3, France
e-mail: jarthur.micoulaud@gmail.com

J. Vion-Dury
e-mail: jean.vion-dury@ap.hm.fr

© Springer-Verlag London 2014

E.R. Miranda and J. Castet (eds.), *Guide to Brain-Computer Music Interfacing*,
DOI 10.1007/978-1-4471-6584-2_4

enabling an expert or an intuitive sound manipulation. Finally, as a perspective to this survey, explicit and implicit brain-computer interfaces (BCI) are described with respect to the previous theoretical frameworks, and a semiotic-based BCI aiming at increasing the intuitiveness of synthesis control processes is envisaged. These interfaces may open for new applications adapted to either handicapped or healthy subjects.

4.1 Introduction

In this article, we present different approaches to sound synthesis and control issues and describe how these procedures can be conceptualized and related to different paradigms within the domain of cognitive neuroscience. A special emphasis is put on the notion of intuitive control and how such a control can be defined from the identification of signal invariants obtained both from the considerations of the physical or signal behaviour of the sound-generating sources and the perceptual impact of the sounds on the listeners.

Since the first sounds were produced by a computer in the late 1950s, computer-based (or synthesized) sounds have become subject to an increasing attention for everyday use. In early years of sound synthesis, the majority of applications were dedicated to musicians who learned to play new instruments that generally offered a lot of control possibilities, but required high skills to operate. Due to increasingly powerful computers, new applications linked to communication, virtual reality and sound design have made sound synthesis available for a broader community. This means that synthesis tools need to be adapted to non-expert users and should offer intuitive control interfaces that do not require specific training. The construction of such intuitive synthesis tools requires knowledge about human perception and cognition in general and how a person attributes sense to sounds. Why are we for instance able to recognize the material of falling objects simply from the sounds they produce, or why do we easily accept the ersatz of horse hooves made by the noise produced when somebody is knocking coconuts together? Is the recognition of sound events linked to the presence of specific acoustic morphologies that can be identified by signal analysis? In the approach presented here, we hypothesize that this is the case and that perception emerges from such invariant sound structures, so-called *invariant sound morphologies*, in line with the ecological approach of visual perception introduced by (Gibson 1986). From a synthesis point of view, this theoretical framework is of great interest, since if enables the conception of perceptually optimized synthesis strategies with intuitive control parameters.

Sound synthesis based on the modelling of physical sources is generally divided in two main classes, i.e. physical models and signal models. Physical models aim at simulating the physical behaviour of sound sources (i.e. the physical origin of sounds), while signal models imitate the recorded signal using mathematical representations without considering the physical phenomena behind the sound production. In the case of physical models, an accurate synthesis can only be achieved when physical phenomena linked to the sound production are well described by

physics. This is not the case for complex sources (e.g. natural phenomena such as wind, rain, fire, etc.). In the case of signal models, any sound can generally be perfectly resynthesized for instance from the analysis of real sounds, independently of the complexity of the underlying physical phenomena of the sound source. However, the control of such sounds is a difficult issue due to the large number of synthesis parameters that generally are implied in such models and to the impossibility to physically interpret these parameters. The physical and signal models can also be combined to form so-called hybrid models (e.g. Ystad and Voinier 2001). The control of these models requires an expertise and the quality judgment of the control is based on an error function linked to the physical or signal precision between the model and the real vibration. Such controls necessitate a scientific expertise apart from certain cases such as musical applications where the control parameters correspond to physical values controlled by the musician (e.g. pressure, force, frequency, etc.). In this latter case, the musical expertise enables the control.

To propose efficient synthesis models that enable intuitive control possibilities, synthesis models combined with perceptual considerations have been developed lately. Perceptual correlates have been sought by testing the perceptual relevance of physical and/or signal parameters through listening tests (cf. Sect. 4.3.2). In the case of environmental sounds, we have identified such perceptually relevant sound morphologies through several experiments. These experiments have made it possible to identify sound elements, also described as sound "atoms", specific to given sound categories that enable definition of high-level control parameters for real-time synthesis applications. Such synthesis tools allow users to synthesize auditory scenes using intuitive rather than reflective processes. Intuitive processes appeal on intuition which is a kind of immediate knowledge, which does not require reasoning, or reflective thought. Intuition can also be defined as the knowledge of an evident truth, a direct and immediate seeing of a thought object (Lalande 1926). The quality of the control strategy is in this case based on perceptual judgments and on easily understandable control parameters on the user interface. Therefore, we call this synthesis control, *intuitive control*.

When searching for perceptually relevant sound morphologies, the understanding of attribution of sense of sounds becomes essential. This issue is a natural part of a more general research field called semiotics that consists in studying the general theory of signs. The notion of signs has been addressed since antiquity by the stoic philosophers (Nadeau 1999). Classically, semiotics is divided in syntax, semantics and pragmatics. Semiology is a part of semiotics, which concerns the social life, and dynamic impact of signs, as language (Nadeau 1999). For de Saussure, language constitutes a special system among all semiological facts. In linguistics, for de Saussure, a sign is the association of a signifier (acoustic image) and a signified (the correlated concept) linked together in a consubstantial way (de Saussure 1955). This consubstantial relationship is often difficult to understand. Semiotics span over both linguistic and non-linguistic domains such as music, vision, biology, etc. This means that it is possible to propose a semiotic approach of sounds, without referring to linguistic semiology. Like in de Saussure construction of signs, one can postulate that every natural (environmental) or social sound is

linked to the afferent concept in the same consubstantial way. For example, if I hear a bell, I immediately know that it is a bell, and perhaps, but not always, I even manage to imagine the size of the bell, depending on its spectral contents. Except for "abstract sounds", i.e., sounds for which the sources cannot be easily identified, one can say that each sound can be considered as a non-linguistic sign whose origin can be described using language, in a reflective thought. Previous studies have shown that the processing of both linguistic and non-linguistic target sounds in conceptual priming tests elicited similar relationships in the congruity processing (cf. Sect. 4.5). These results indicate that it should be possible to draw up a real semiotic system of sounds, which is not the linguistic semiology, because phonemes can be considered only as particular cases of sounds.

So far, the identification of signal invariants has made it possible to propose an intuitive control of environmental sounds from verbal labels or gestures. An interesting challenge in future studies would be to propose an even more intuitive control of sound synthesis processes that bypasses words and gestures and directly uses a BCI that records electroencephalographic signals in a BCI/synthesizer loop. This idea is not new and several attempts have already been made to pilot sounds directly from the brain activity. In (Väljamäe et al. 2013), the authors made an exhaustive review in the field of EEG sonification in various applications (medical, neurofeedback, music, etc.) and concluded that the type of mapping strategy strongly depends on the applications. For instance, in the case of musical applications, the mapping is generally determined by artistic choices and does not necessarily mirror a strict semiotic relation. The intuitive BCI-controlled synthesizer that we aim at is intended for a generic context and should enable the identification of brain activity linked to specific signal morphologies that reflect the attribution of sense to a sound.

This paper is organized as follows. In Sect. 4.2, the methodology that leads to intuitive sound synthesis is viewed in the light of representational-computational and enactive perspectives. Then, in Sect. 4.3, two sound synthesis approaches are described and related to the previously presented perspectives. In Sect. 4.4, different control strategies emanating from the different synthesis approaches are described. In Sect. 4.5, some results from experiments supporting the existence of semiotics for non-linguistic sounds are presented. Finally, in Sect. 4.6, a prospective view on a control strategy for synthesis processes based on a BCI is proposed.

4.2 Two Conceptions on the Way We Interact with the Surrounding World

Sound synthesis that integrates perceptual effects from the morphology of their signal in order to enable intuitive control to the end-user brings forward the following questions: How do I attribute a meaning to a perceived sound (related to the semiotics)? What effect does this sound have on me? These questions induce a change in our position with respect to the sound from a third-person position

(observer) in more traditional synthesis approaches where only acoustic considerations are taken into account, to a first-person position (implied) in the perceptual synthesis processes. This corresponds to a change from a representational to a neurophenomenological point of view in the field of cognitive neuroscience (Varela 1996). We here adopt a similar change of viewpoint to investigate the phenomenon of sound perception as it was seminally studied in (Petitmengin et al. 2009).

Classically, in the standard paradigm of cognitive neuroscience, there is, on one hand, the physical object and on the other hand, the subject that perceives this object according to his/her mental representation of the physical reality. From this conception of representation proposed by Descartes, a representational-computational paradigm has been developed. This paradigm involves the existence of a *correct* representation of the physical world and assumes that the perception of the object is all the more adequate when the subject's mental representation matches the physical reality, considered as the reference (Varela 1989). Less classically, in the neurophenomenological paradigm of cognitive sciences, it is the interaction between the subject and the object, which enables the subject to perceive an object. F. Varela called this interaction: *enaction* (Varela 1989; Varela et al. 1991). In the enactive paradigm, the mind and the surrounding world are mutually imbricated. This conception is inspired from the phenomenological philosophy of Husserl, who called this interaction a noetic–noematic correlation (Husserl 1950). He posited that there was a link between intentional content on the one hand, and extra-mental reality on the other, such that the structure of intentionality of the consciousness informs us about how we perceive the world as containing particular objects. In a certain manner, and quite caricatured, the physical reality is no more the reference, and the subject becomes the reference. The perception of the object is all the more adequate when the subject's perception makes it possible to efficiently conduct an action to respond to a task. As Varela puts it (Varela et al. 1991):

> The enactive approach underscores the importance of two interrelated points: 1) perception consists of perceptually guided action and 2) cognitive structures emerge from the recurrent sensorimotor patterns that enable action to be perceptually guided.

and concludes:

> We found a world enacted by our history of structural coupling.

In 1966, P. Schaeffer, who was both a musician and a researcher, published the "Traité des objets musicaux" (Schaeffer 1966), in which he reported more than ten years of research on electroacoustic music. He conducted a substantial work that was of importance for electroacoustic musicians. With a multidisciplinary approach, he intended to carry out fundamental music research that included both Concrete[1] and traditional music. Interestingly, he naturally and implicitly adopted a phenomenological approach to investigate the sound perception in listening

[1] The term "concrete" is related to a compositional method which is based on concrete material, i.e., recorded or synthesized sounds, in opposition with "abstract" music which is composed in an abstract manner, i.e., from ideas written on a score, and becomes "concrete" afterwards.

experiences. In particular, he introduced the notion of *sound object*. The proposition of P. Schaeffer naturally conducts the acoustician from the representational-computational paradigm to the enactive paradigm, since P. Schaeffer in line with the phenomenological viewpoint stresses the fact that sound perception is not only related to a correct representation of the acoustic signal. This is also coherent with later works of Varela and the conception of perception as an enactive process, where the sound and the listener constitute a unique imbricated system. The perception of the sound is modified by the intentionality of the subject directed towards the sound, which can induce an everyday listening, which is a source-oriented kind of listening, or musical (or acousmatic) listening, which involves the perception of the quality of the sound (Gaver 1993a, b). Thus, sound synthesis should not be limited to the simulation of the physical behaviour of the sound source. In other words, it is the sound object given in the process of perception that determines the signal to be studied, meaning that perception has to be taken into account during the signal reconstruction process.

In the work of P. Schaeffer, morphology and typology have been introduced as analysis and creation tools for composers as an attempt to construct a music notation that includes electroacoustic music and therefore any sound. This typological classification is based on a characterization of spectral (mass) and dynamical (facture) profiles with respect to their complexity and consists of 28 categories. There are nine central categories of "balanced" sounds for which the variations are neither too rapid and random nor too slow or non existent. Those nine categories include three facture profiles (sustained, impulsive or iterative) and three mass profiles (tonic, complex and varying). On both sides of the balanced objects in the table, there are 19 additional categories for which mass and facture profiles are very simple/repetitive or vary a lot. This classification reveals perceptually relevant sound morphologies and constitute a foundation for studies on intuitive sound synthesis.

Based on these previous theoretical frameworks from cognitive neuroscience, we suggest that the control of sound synthesis can be discussed in the framework of the representational-computational and the enactive points of view. In the approach inspired by the representational-computational framework, we consider that the user controls physical or signal parameters of the sound with the idea that the more actual (with respect to the physical reality) the parameter control, the better the perception. The physical or signal properties of sounds are considered as the reference for the sound control. In the approach inspired by the enactive framework, we consider that the user is involved in an interactive process where he/she controls the sound guided by the perceptual effect of his/her action. The idea is that the more recurrent (and intuitive) the sensorimotor manipulation, the better the perception. The sound control enables the perception to become a perceptually guided action. This is an enactive process because the sound influences the control effectuated by the subject and the control action modifies the sound perception. The sound as perceived by the subject is thus the reference for the sound control. Such enactive framework formed a theoretical basis for a recent research community centred on the conception of new human–computer interfaces (Enactive Network) and in a

natural way, led to numerous interactive applications in musical contexts (*Journal of New Music Research, special issue* "*Enaction and Music*" 2009). A general review on fundamental research in the field of enactive music cognition can be found in (Matyja and Schiavio 2013).

4.3 Sound Synthesis Processes

To date, two approaches to synthesize sounds could be highlighted: sound synthesis based on the modelling of physical sources (from either physical or signal perspectives) and sound synthesis based on the modelling of perceptual effects. Interestingly, these synthesis approaches could be linked to the two paradigms related to our perception of the surrounding world (i.e. approaches inspired by the representational-computational and the enactive paradigms, cf. Fig. 4.2) described in the previous section.

4.3.1 Two Approaches for Sound Synthesis

4.3.1.1 Modelling the Physical Sources

In the case of sound synthesis based on the modelling of physical/vibrating sources, either the mechanical behaviour or the resulting vibration of the sound source is simulated.

Physical synthesis models that simulate the physical behaviour of sound sources can either be constructed from the equations describing the behaviour of the waves propagating in the structure and their radiation in air (Chaigne 1995) or from the behaviour of the solution of the same equations (Karjalainen et al. 1991; Cook 1992; Smith 1992; Bilbao 2009). Physical models have been used to simulate a large number of sound sources from voice signals to musical instruments. Several synthesis platforms based on physical modelling are now available, such as Modalys that is based on modal theory of vibrating structures that enable the simulation of elementary physical objects such as strings, plates, tubes, etc. These structures can further be combined to create more complex virtual instruments (http://forumnet.ircam.fr/product/modalys/?lang=en)n.d). Cordis-Anima is a modelling language that enables the conception and description of the dynamic behaviour of physical objects based on mass-spring-damper networks (http://www-acroe.imag.fr/produits/logiciel/cordis/cordis_en.htmln.d). Synthesis models for continuous interaction sounds (rolling, scratching, rubbing, etc.) were proposed in previous studies. In particular, models based on physical modelling or physically informed considerations of such sounds can be found (Gaver 1993a; Hermes 1998; van den Doel et al. 2001; Pai et al. 2001; Rath and Rocchesso 2004; Stoelinga and Chaigne 2007). In particular, Avanzini et al. (2005) developed a physically based synthesis model for friction sounds. This model generates realistic sounds of continuous contact between rubbed surfaces (friction, squeaks, squeals, etc.). The

parameters of the model are the tribological properties of the contact condition (stiffness, dynamic or static friction coefficients, etc.) and the dynamic parameters of the interaction (mainly the velocity and the normal force). Also, a synthesis technique based on the modal analysis of physical objects (finite element modelling of each object for precomputation of shapes and frequencies of the modes) was proposed by (O'Brien et al. 2002) in the context of interactive applications. Note that this approach presents a limitation when the physical considerations involve complex modelling and can less easily be taken into account for synthesis perspectives especially with interactive constraints.

Signal synthesis models that simulate the resulting vibration of the sound source are based on a mathematical modelling of the signal. They are numerically easy to implement and can be classified in three groups as follows:

- Additive synthesis: The sound is constructed as a superposition of elementary sounds, generally sinusoidal signals modulated in amplitude and frequency (Risset 1965). For periodic or quasi-periodic sounds, these components have average frequencies that are multiples of one fundamental frequency and are called harmonics. The amplitude and frequency modulation (FM) laws should be precise when one reproduces a real sound. The advantage of these methods is the potential for intimate and dynamic modifications of the sound. Granular synthesis can be considered as a special kind of additive synthesis, since it also consists in summing elementary signals (grains) localized in both the time and the frequency domains (Roads 1978).
- Subtractive synthesis: The sound is generated by removing undesired components from a complex sound such as noise. This technique is linked to the theory of digital filtering (Rabiner and Gold 1975) and can be related to some physical sound generation systems such as speech (Flanagan et al. 1970; Atal and Hanauer 1971). The advantage of this approach is the possibility of uncoupling the excitation source and the resonance system. The sound transformations related to these methods often use this property to make hybrid sounds or crossed synthesis of two different sounds by combining the excitation source of a sound and the resonant system of another (Makhoul 1975; Kronland-Martinet 1989).
- Global (or non-linear) synthesis: The most well-known example of such methods is audio FM. This technique updated by Chowning (1973) revolutionized commercial synthesizers. The advantages of this method are that it calls for very few parameters, and that a small number of numerical operations can generate complex spectra. They are, however, not adapted to precise signal control, since slight parameter changes induce radical signal transformations. Other related methods such as waveshaping techniques (Arfib 1979; Le Brun 1979) have also been developed.

In some cases, both approaches (physical and signal) can be combined to propose hybrid models, which have shown to be very useful when simulating certain musical instruments (Ystad and Voinier 2001; Bensa et al. 2004).

4.3.1.2 Modelling the Perceptual Effects

In the case of sound synthesis based on the modelling of perceptual effects, the sound generation is not merely based on the simulation of the physical or signal phenomena. This approach enables the synthesis of any kind of sounds, but it necessitates the understanding of the perceptual relevance of the sound attributes that characterize the sound category in question. Concerning environmental sounds, several studies have dealt with the identification and classification of such sounds (Ballas 1993; Gygi and Shafiro 2007; Gygi et al. 2007; Vanderveer 1979). A hierarchical taxonomy of everyday sounds was proposed by Gaver (1993b) and is based on three main categories: sounds produced by vibrating solids (impacts, deformation, etc.), aerodynamic sounds (wind, fire, etc.) and liquid sounds (drops, splashes, etc.). This classification related with the physics of sound events and has shown to be perceptually relevant. Hence, the perceptual relevance of these categories encourages the search for invariant sound morphologies specific to each category. This notion is developed in the next section.

4.3.2 Invariant Sound Morphologies

The invariant sound morphologies associated with the evocation of sound attributes can either be linked to the physical behaviour of the source (Giordano and McAdams 2006), to the signal parameters (Kronland-Martinet et al. 1997) or to timbre qualities based on perceptual considerations (McAdams 1999). This means that different synthesis approaches can be closely related, since in some cases, physical considerations and in other cases, signal variations might reveal important properties to identify the perceived effects of the generated sounds. In particular for environmental sounds, several links between the physical characteristics of actions (impact, bouncing, etc.), objects (material, shape, size, cavity, etc.) and their perceptual correlates were established in previous studies (see Aramaki et al. 2009; Aramaki et al. 2011 for a review). In summary, the question of sound event recognition was subject to several inquiries (e.g. Warren and Verbrugge 1984; Gaver 1993a, b) inspired by Gibson's ecological approach (Gibson 1986) and latter formalized by McAdams and Bigand (1993). This led to the definition of structural and transformational invariants linked to the recognition of the object's properties and its interaction with the environment, respectively.

Sounds from impacted objects: Impact sounds have been largely investigated in the literature from both physical and perceptual points of view. Several studies revealed relationships between perceptual attributes of sound sources and acoustic characteristics of the produced sound. For instance, the attack time has been related to the perception of the hardness of the mallet that was used to impact the resonant object, while the distribution of the spectral components (described by inharmonicity or roughness) of the produced sound has been related to the perceived shape of the object. The perceived size of the object is mainly based on the pitch. A physical explanation can be found in the fact that large objects vibrate at lower eigenfrequencies than small ones. Finally, the perception of material seems to be

linked to the damping of the sound that is generally frequency-dependent: high frequency components are damped more heavily than low frequency components. In addition to the damping, the density of spectral components, which is directly linked to the perceived roughness, was also shown to be relevant for the distinction between metal versus glass and wood categories (Aramaki et al. 2009, 2011).

Sounds from continuous interactions: Based on previous works described in Sect. 4.3, invariant sound morphologies related to the perception of interactions such as rubbing, scratching and rolling were investigated (Conan et al. 2013a, b; Thoret et al. 2013). An efficient synthesis model, init. An efficient synthesis model, initially proposed by (Gaver 1993a) and improved by (van den Doel et al. 2001), consists in synthesizing the interaction sounds by a series of impacts that simulates the successive micro-impacts between a plectrum and the asperities of the object's surface. Therefore, it has been highlighted that a relevant sound invariant morphology allowing the discrimination between rubbing and scratching interactions was the temporal density of these impacts, i.e., the more (respectively, the less) impacts that occur, the more the sound is associated to rubbing (respectively, to scratching) (Conan et al. 2012). For the rolling interaction, it has been observed, from numerical simulations based on a physical model, that the temporal structure of the generated impact series follows a specific pattern. In particular, the time intervals between impacts and associated amplitudes are strongly correlated. Another fundamental aspect supported by physical considerations is the fact that the contact time of the impact depends on the impact velocity. This dependency also seems to be an important auditory cue responsible for the evocation of a rolling interaction (Conan et al. 2013).

These studies related to such interaction sounds led us to address the perceptual relation between the sound and the underlying gesture that was made to produce the sound. Many works highlighted the importance of the velocity profile in the production of a movement and its processing may be involved at different levels of perception of a biological movement both in the visual and in the kinaesthetic system ((Viviani and Stucchi 1992; Viviani et al. 1997; Viviani 2002) for a review). Based on these findings, we investigated whether the velocity profile, in the case of graphical movements, was also a relevant cue to identify a human gesture (and beyond the gesture, the drawn shape) from a friction sound. Results from a series of perceptual experiments revealed that the velocity profile transmits relevant information about the gesture and the geometry of the drawn shape to a certain extent. Results also indicated the relevance of the so-called 1/3-power law, defined from seminal works by Viviani and his colleagues and translating a biomechanics constraint between the velocity of a gesture and the local curvature of the drawn shape, to evoke a fluid and natural human gesture through a friction sound (cf. Thoret et al. 2013, 2014 for details and review).

Other environmental sounds: For other classes of environmental sounds such as wave or aerodynamic sounds, physical considerations generally involve complex modelling and signal models are then useful. From a perceptual point of view, these sounds evoke a wide range of different physical sources, but interestingly, from a signal point of view, common acoustic morphologies can be highlighted across these sounds. We analysed several signals representative of the main categories of

environmental sounds as defined by Gaver and we identified a certain number of perceptually relevant signal morphologies linked to these categories (Gaver 1993a, b). To date, we concluded on five elementary sound morphologies based on impacts, chirps and noise structures (Verron et al. 2009). This finding is based on a heuristic approach that has been verified on a large set of environmental sounds. Granular synthesis processes based on these five sound "atoms" then enabled the generation of various environmental sounds (i.e. solid interactions, aerodynamic or liquid sounds). Note that this atom dictionary may be completed or refined in the future without compromising the proposed methodology.

A first type of grain is the "tonal solid grain" that is defined by a sum of exponentially damped sinusoids. Such a grain is well adapted to simulate sounds produced by solid interactions. Nevertheless, this type of grain cannot alone account for any kind of solid impact sounds. Actually, impact sounds characterized by a strong density of modes or by a heavy damping may rather be modelled as an exponentially damped noise. This sound characterization stands for both perceptual and signal points of view, since no obvious pitch can be extracted from such sounds. Exponentially damped noise constitutes the second type of grain, the so-called "noisy impact grain". Such a grain is well adapted to simulate crackling sounds. The third type of grain concerns liquid sounds. From an acoustic point of view, cavitation phenomena (e.g. a bubble in a liquid) lead to local pressure variations that generate time-varying frequency components such as exponentially damped linear chirps. Hence, the so-called "liquid grain" consists of an exponentially damped chirp signal. Finally, aerodynamic sounds generally result from complicated interactions between solids and gases and it is therefore difficult to extract useful information from corresponding physical models. A heuristic approach allowed us to define two kinds of aerodynamic grains: the "whistling grain" (slowly varying narrow band noise) and the "background aerodynamic grain" (broadband filtered noise). Such grains are well adapted to simulate wind and waves.

By combining these five grains using an accurate statistics of appearance, various environmental auditory scenes can be designed such as rainy ambiances, seacoast ambiances, windy environments, fire noises, or solid interactions simulating solid impacts or footstep noises. We currently aim at extracting the parameters corresponding to these grains from the analysis of natural sounds, using matching pursuit like methods. Perceptual evaluations of these grains will further allow us to identify or validate signal morphologies conveying relevant information on the perceived properties of the sound source.

4.4 Control Strategies for Synthesis Processes

The choice of synthesis model highly influences the control strategy. Physical synthesis models have physically meaningful parameters, which might facilitate the interpretation of the consequence of the control on the resulting sound. This is less so for signal models obtained from mathematical representations of sounds.

Perceptual considerations might, however, be combined with these models to propose intuitive control strategies as described in the following sections.

4.4.1 Control of Synthesis Parameters

Although physical models can produce high-quality sounds that are useful for instance for musical purposes, this approach is less adapted to environmental sounds, when the physics of such sound sources is not sufficiently well understood or the existing models are not real-time compatible. In such cases, signal models that enable the simulation of the sound vibrations through mathematical models are useful. The control of these models consists in manipulating physical or signal parameters. Practically, these approaches might involve the control of physical variables (for instance, characterizing the tribological or mechanical properties of the source) or a high number of synthesis parameters (up to a hundred) that are generally not intuitive for a non-expert user. This means that a certain scientific (or musical) expertise is needed to use such models (expert control). In fact, the calibration of the control of these models is based on an *error function* that reveals the difference between the model and the actual physical sound vibration (cf. Fig. 4.2).

4.4.2 Control of Perceptual Effects

Common to all the previous approaches described in Sect. 4.4.1 is the lack of perceptual criteria. Actually, since the timbre of the resulting sound is generally related to the synthesis parameters in a non-linear way, the control process can quickly become complicated and non-intuitive. The design of a control of perceptual effects may lead to the definition of an intuitive control strategy. In particular, based on the identification of invariant sound morphologies (cf. Sect. 4.3.2), control processes mediating various perceptual evocations could be designed. In line with the previous definitions of structural and transformational invariants, the framework of our control strategy is based on the so-called {*action/object*} paradigm, assuming that the produced sound can be defined as the consequence of an action on an object. This approach supports the determination of sound morphologies that carry information about the action and the object, respectively.

Here we present several synthesis tools that we have developed for generating and intuitively controlling sounds. These synthesis models make it possible to relevantly resynthesize natural sounds. In practice, we adopted hierarchical levels of control to route and dispatch the parameters from an intuitive to the algorithmic level. As these parameters are not independent and might be linked to several signal properties at a time, the mapping between levels is far from being straightforward.

Sounds from impacted objects: We have developed an impact sound synthesizer offering an intuitive control strategy based on a three-level architecture (Aramaki et al. 2010a) (cf. Fig. 4.1). The top layer gives the user the possibility to define the impacted object using verbal descriptions of the object (nature of the

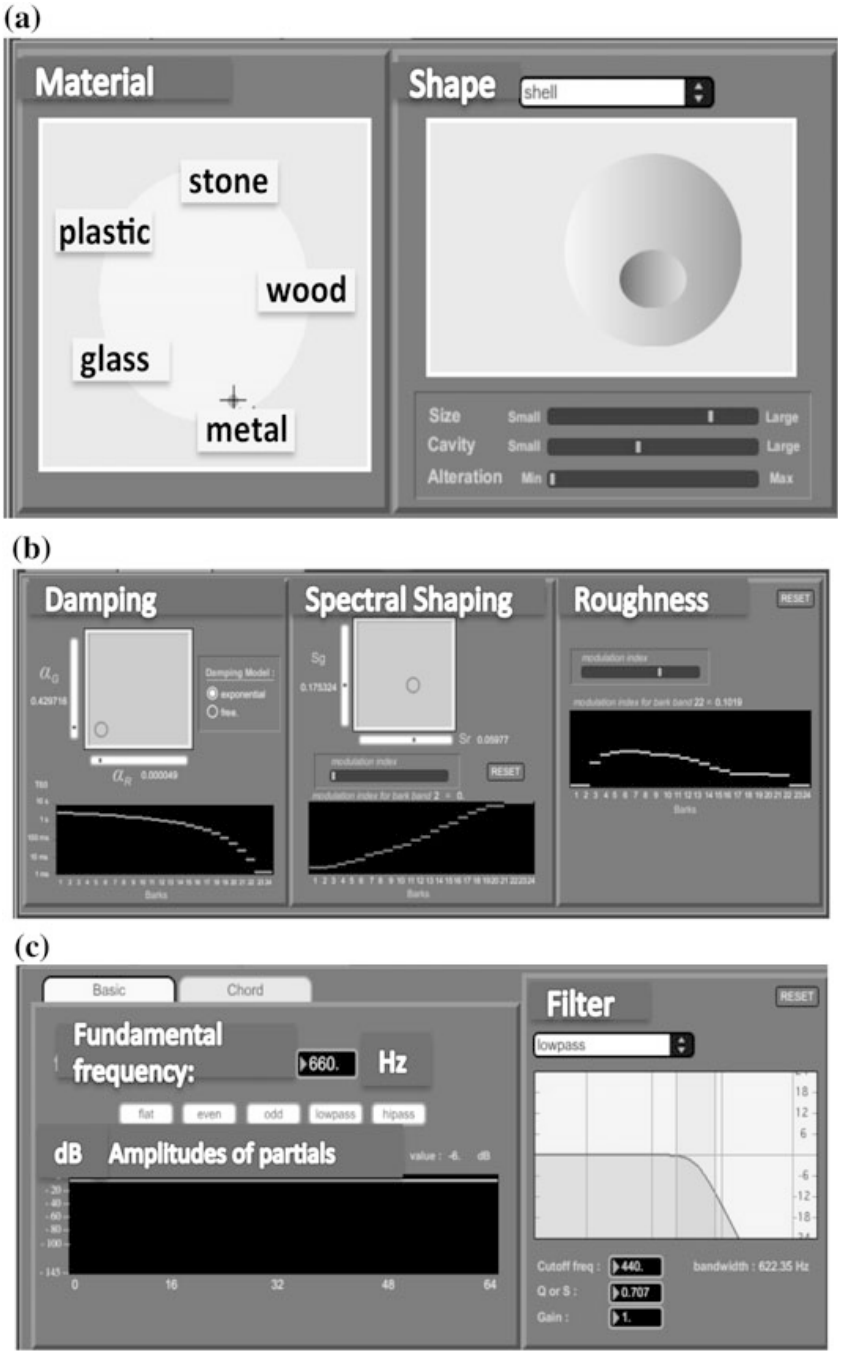


Fig. 4.1 **a** *Top layer* (semantic labels describing the perceived material and shape of the object), **b** *middle layer* (acoustic descriptors) and **c** *bottom layer* (synthesis parameters of the signal model) designed for the control of the impact sound synthesizer

perceived material, size and shape, etc.) and the excitation (force, hardness of the mallet, impact position, etc.). The middle layer is based on perceptually relevant acoustic descriptors linked to the invariant sound morphologies (cf. Sect. 4.3.2). The bottom layer consists of the set of synthesis parameters (for expert users). Two mapping strategies are implemented between the layers (we refer to (Aramaki et al. 2010a) for more details). The first one focuses on the relationships between verbal

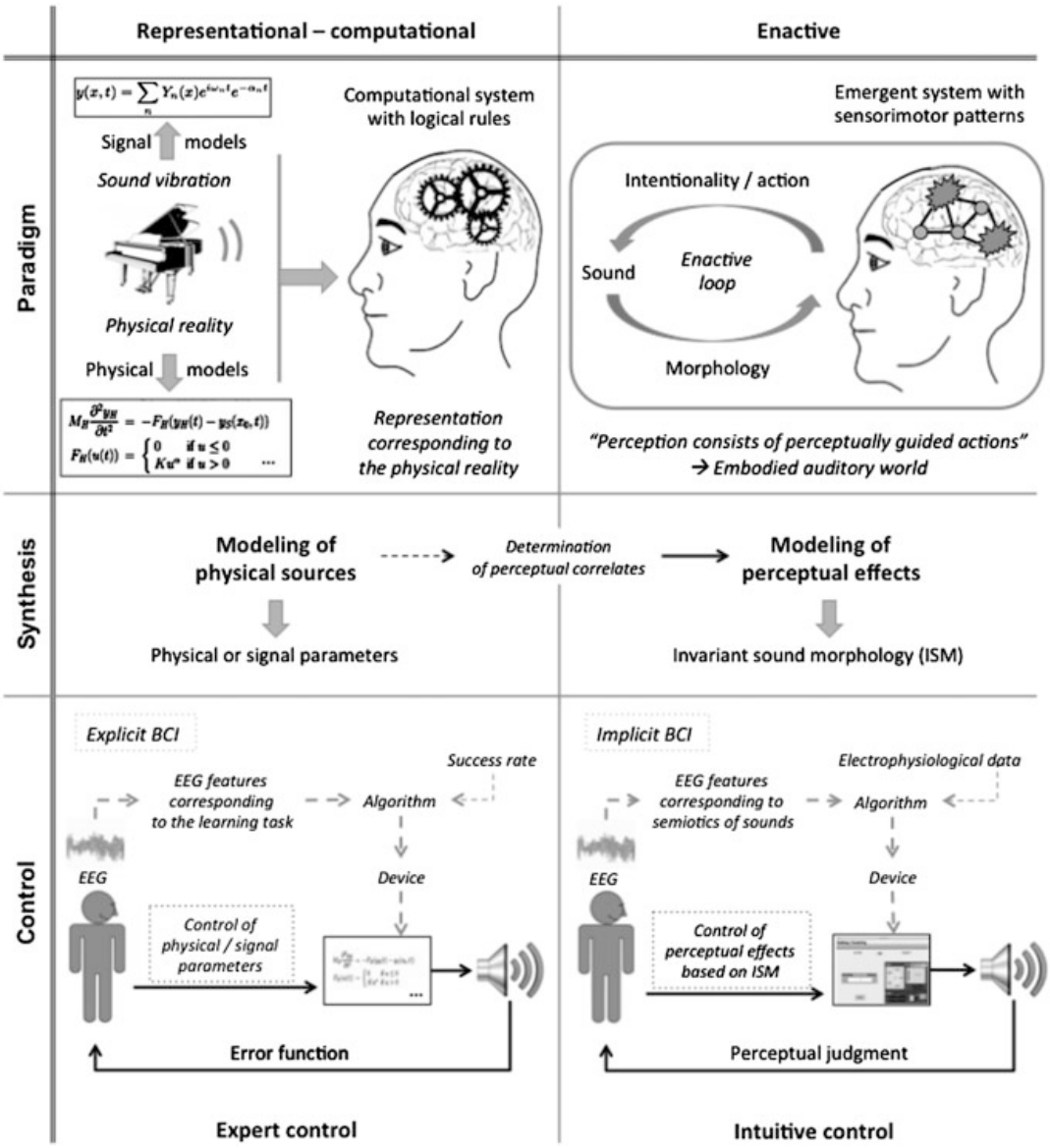


Fig. 4.2 General didactic synopsis including two approaches inspired by the representational-computational and enactive paradigms from cognitive neuroscience, the associated viewpoints for sound synthesis (modelling of physical sources and modelling of perceptual effects) and sound control (expert and intuitive control). A prospective view on the use of BCI in the context of sound synthesis control is also illustrated

descriptions of the sound source and the sound descriptors (damping, inharmonicity, roughness, etc.) characterizing perceptually relevant sound morphologies. The second one focuses on the relationships between sound descriptors and synthesis parameters (damping coefficient, amplitude and frequency of the components).

Sounds from continuous interactions: Control strategies for the synthesis processes of such sounds have recently been developed. In particular, an intuitive control strategy adapted to a non-linear friction sound model (producing phenomena such a creaky door, a singing glass or a squeaking wet plate) has been

proposed. Inspired from Schelleng's diagrams, the proposed control is defined from a flexible physically informed mapping between a dynamic descriptor (velocity, pressure), and the synthesis parameters, and allows coherent transitions between the different non-linear friction situations (Thoret et al. 2013). Another intuitive control strategy dedicated to rolling sound synthesis has also been proposed (Conan et al. 2013). This strategy is based on a hierarchical architecture similar to that of the impacted object sounds (cf. previous paragraph). The high-level controls that can be manipulated by the end-user are the characteristics of the rolling ball (i.e. size, asymmetry and speed) and the irregularity of the surface. The low-level parameters (e.g. impacts' statistics, modulation frequency and modulation depth) are modified accordingly with respect to the defined mapping. Recently, a control strategy enabling to perceptually morph between the three continuous interactions, i.e. rubbing, scratching and rolling, was designed. For that purpose, we developed a synthesis process that is generic enough to simulate these different interactions and based on the related invariant sound morphologies (cf. Sect. 4.3.2). Then, a perceptual "interaction space" and the associated intuitive navigation strategy were defined with given sound prototypes considered as anchors in this space (Conan et al. 2013).

Finally, in line with the *action/object* paradigm, the complete synthesis process has been implemented as a source-filter model. The resulting sound is then obtained by convolving the excitation signal (related to the nature of the interaction) with the impulse response of the resonating object. The impulse response is implemented as a resonant filter bank, which central frequencies correspond to the modal frequencies of the object.

Immersive auditory scenes: An intuitive control of the sound synthesizer dedicated to environmental auditory scenes was defined. The control enables the design of complex auditory scenes and included the location and the spatial extension of each sound source in a 3D space so as to increase the realism and the feeling of being immersed in virtual scenes. This control is particularly relevant to simulate sound sources such as wind or rain that are naturally diffuse and wide. In contrast with the classical two-stage approach, which consists in first synthesizing a monophonic sound (timbre properties) and then spatializing the sound (spatial position and extension in a 3D space), the architecture of the proposed synthesizer yielded control strategies based on the overall manipulation of timbre and spatial attributes of sound sources at the same level of sound generation (Verron et al. 2010).

The overall control of the environmental scene synthesizer can be effectuated through a graphical interface where the sound sources (selected among a set of available sources: fire, wind, rain, wave, chimes, footsteps, etc.) can be placed around the listener (positioned in the centre of the scene) by defining the distance and the spatial width of each source. The sources are built from the elementary grains defined previously in Sect. 4.3.2. A fire scene is for instance built from a combination of a whistling grain (simulating the hissing), a background aerodynamic grain (simulating the background combustion) and noisy impact grains (simulating the cracklings). The latter grains are generated and launched randomly

with respect to time using an accurate statistic law that can be controlled. A global control of the fire intensity, mapped with the control of the grain generation (amplitude and statistic law), is then designed. A rainy weather sound ambiance can be designed with a rain shower, water flow and drops, each of these environmental sounds being independently spatialized and constructed from a combination of the previous grains (see Verron et al. 2009 for more details). In case of interactive uses, controls can be achieved using either MIDI interfaces, from data obtained from a graphical engine or other external data sources.

4.5 Evidence of Semiotics for Non-linguistic Sounds

To propose an even more intuitive control of sound synthesis that directly uses a BCI, a relationship between the electroencephalogram (EEG) and the nature of the underlying cerebral processes has to be investigated. We here present results of several experimental studies aiming at supporting the existence of semiotics for non-linguistic sounds. In these studies, we used either synthetic stimuli using analysis/transformation/synthesis processes or sounds of a specific kind called "abstract" sounds promoting acousmatic listening (cf. Sect. 4.2). The participants' responses and reaction times (RTs) provided objective measurements to the processing of stimulus complexity.

Electrophysiological data: When appropriate, we also investigated the neural bases of the involved brain processes by analysing the EEG with the method of event-related potentials (ERP) time-locked to the stimulus onset during the various information processing stages. The ERP elicited by a stimulus (a sound, a light, etc.) are characterized by a series of positive (P) and negative (N) deflections relative to a baseline. These deflections (called components) are defined in terms of their polarity, their maximum latency (relative to the stimulus onset), their distribution among several electrodes placed in standard positions on the scalp and by their functional significance. Components P100, N100 and P200 are consistently activated in response to the auditory stimuli (Rugg and Coles 1995). Several late ERP components (N200, P300, N400, etc.) are subsequently elicited and associated with specific brain processes depending on the experimental design and the task in hand.

4.5.1 Perceptual Categorization of Sounds from Impacted Materials

In this experiment, we studied the perception of sounds obtained from impacted materials, in particular, wood, metal and glass (Aramaki et al. 2010a; Aramaki et al. 2010b; Aramaki et al. 2011). For this purpose, natural sounds were recorded, analysed, resynthesized and tuned to the same chroma to obtain sets of synthetic sounds representative of each category of the selected material. A sound-morphing process (based on an interpolation method) was further applied to obtain sound continua simulating progressive transitions between materials. Although sounds

located at the extreme positions on the continua were indeed perceived as typical exemplars of their respective material categories, sounds in intermediate positions, which were synthesized by interpolating the acoustic parameters characterizing sounds at extreme positions, were consequently expected to be perceived as ambiguous (e.g. to be neither wood nor metal). Participants were asked to categorize each of the randomly presented sounds as wood, metal or glass.

Based on the classification rates, we defined "typical" sounds as sounds that were classified by more than 70 % of the participants in the right material category and "ambiguous" sounds, those that were classified by less than 70 % of the participants in a given category. Ambiguous sounds were associated with slower RTs than typical sounds. As might be expected, ambiguous sounds are therefore more difficult to categorize than typical sounds. This result is in line with previous findings in the literature showing that non-meaningful sounds were associated with longer RTs than meaningful sounds. Electrophysiological data showed that ambiguous sounds elicited more negative ERP (a negative component, N280, followed by a negative slow wave, NSW) in fronto-central brain regions and less positive ERP (P300 component) in parietal regions than typical sounds. This difference may reflect the difficulty to access information from long-term memory. In addition, electrophysiological data showed that the processing of typical metal sounds differed significantly from those of typical glass and wood sounds as early as 150 ms after the sound onset. The results of the acoustic and electrophysiological analyses suggested that spectral complexity and sound duration are relevant cues explaining this early differentiation. Lastly, it is worth noting that no significant differences were observed on the P100 and N100 components. These components are known to be sensitive to sound onset and temporal envelope, reflecting the fact that the categorization process occurs in later sound-processing stages.

4.5.2 Conceptual Priming for Non-linguistic Sounds

In language, a comprehensible linguistic message is for instance conveyed by associating words while respecting the rules of syntax and grammar. Can similar links be generated between non-linguistic sounds so that any variation will change the global information conveyed? From the cognitive neuroscience point of view, one of the major issues that arises from this question is whether similar neural networks are involved in the allocation of meaning in the case of language and that of sounds of other kinds. In a seminal study using a priming procedure, Kutas and Hillyard (Kutas and Hillyard 1980) established that the amplitude of a negative ERP component, the N400 component, increases when final sentence words are incongruous (e.g. *The fish is swimming in the river/carpet*). Since then, the N400 has been widely used to study semantic processing in language. In recent studies, priming procedures with non-linguistic stimuli such as pictures, odours, music and environmental sounds have been used (e.g. Holcomb and McPherson 1994; Castle et al. 2000; Koelsch et al. 2004; Daltrozzo and Schön 2009; Van Petten and Rheinfelder 1995; Orgs et al. 2006). Although the results of these experiments

mostly have been interpreted as reflecting some kind of conceptual priming between words and non-linguistic stimuli, they may also reflect linguistically mediated effects. For instance, watching a picture of a bird or listening to a birdsong might automatically activate the verbal label "bird". Therefore, the conceptual priming cannot be taken to be purely non-linguistic because of the implicit naming induced by the processing of the stimulus. Such conceptual priming might imply at least language, generation of auditory scenes, and mental imaging, at various associative (non specific) cortex area levels. This might probably activate large neural/glial networks using long-distance synchronies, which could be investigated by a synchronous EEG activity measurement (Lachaux et al. 1999).

The aim of our first conceptual priming study (Schön et al. 2010) was to attempt to reduce as far as possible the likelihood that a labelling process of this kind takes place. To this end, we worked with a specific class of sounds called "abstract sounds", which physical sources cannot be easily recognized, meaning that verbal labelling is less likely to take place (Merer et al. 2011). We then conducted conceptual priming tests using word/sound pairs with different levels of congruence between the prime and the target. Subjects had to decide whether or not the prime and the target matched. In the first experiment, a written word was presented visually before the abstract sound, and in the second experiment, the order of presentation was reversed. Results showed that participants were able to assess the relationship between the prime and the target in both presentation orders (sound/word vs. word/sound), showing low inter-subject variability and good consistency. The presentation of a word reduced the variability of the interpretations of the abstract sound and led to a consensus between subjects in spite of the fact that the sound sources were not easily recognizable. Electrophysiological data showed the occurrence of an enhanced negativity in the 250–600-ms latency range in response to unrelated as compared to related targets in both experiments and the presence of a more fronto-central distribution in response to word targets and a more centro-parietal distribution in response to sound targets.

In a subsequent study (Aramaki et al. 2010b), we avoided the use of words as primes or targets. Conceptual priming was therefore studied using impact sounds (also used in the categorization experiment previously presented), as both primes and targets. As described in Sect. 4.5.1, these impact sounds were qualified as either typical or ambiguous with respect to a material category depending on their score in the categorization experiment. 3° of congruence were investigated through various combinations of typical and ambiguous sounds as prime and target: related, ambiguous and unrelated. The priming effects induced in these conditions were compared with those observed with linguistic sounds (spoken words) in the same group of participants. Results showed that N400-like components were also activated in a sound–sound design. This component may therefore reflect a search for meaning that is not restricted to linguistic meaning. Moreover, ambiguous targets also elicited larger N400-like components than related targets for both linguistic and non-linguistic sounds. These findings showed the existence of similar relationships in the processing of semiotics of both non-linguistic and linguistic target sounds. This study clearly means that it is possible to draw up a real language for non-linguistic sounds.

4.6 Towards a Semiotic-Based Brain Computer Interface (BCI)

BCIs provide a link between a user and an external electronic device through his or her brain activity, independently of the voluntary muscle activity of the subject. Most often BCIs are based on EEG recordings that allow for non-invasive measurements of electrical brain activity. As substitutional devices, BCIs open interesting perspectives for rehabilitation, reducing disability and improving the quality of life of patients with severe neuromuscular disorders such as amyotrophic lateral sclerosis or spinal cord injury (Wolpaw et al. 2002). Such interfaces, among many other possibilities, enable patients to control a cursor, to select a letter on a computer screen, or to drive a wheelchair. In addition to medical and substitutional applications, BCIs as enhancing devices can be used with healthy subjects. For example, in the field of video games, BCIs could capture the cognitive or emotional state of the user through the EEG to develop more adaptive games and to increase the realism of the gaming experience (Nijholt 2009). To date, two approaches to BCI could be highlighted: "explicit (or active) BCI" and "implicit (or passive) BCI" (George and Lécuyer 2010). These two classes of BCI could be linked with the two approaches inspired from the paradigms of cognitive science (described in Sect. 4.2) and the two approaches for sound synthesis (described in Sect. 4.3).

4.6.1 Explicit BCI

The explicit BCI is based on the principles of *operant conditioning*, the basic learning concept in experimental psychology, which assumes that the probability of occurrences of an animal or human behaviour is a function of a positive or negative reinforcement during the subject's learning process (Micoulaud-Franchi et al. 2013). Thus, the explicit BCI requires a learning period (George and Lécuyer 2010). In practice, the subject intentionally tries to control his/her cognitive activity to change his/her EEG activity and control an external electronic device. The EEG signal is recorded, processed in real time to extract the information of interest (e.g. spectral power EEG, slow cortical potential or ERP). This information is related to a cognitive activity that the subject intentionally produces. This information is further transmitted to the external electronic device using specific mapping that leads to the control of the device in the desired direction. The positive reinforcement (and the success rate) is determined by the capacity of controlling the external electronic device to achieve a given task.

This configuration fits with traditional neurofeedback therapeutics where the subject learns to intentionally control EEG through visual or auditory positive reinforcement, without any control of external device (Micoulaud-Franchi et al. 2013). In this context, the positive reinforcement could be an increase of a number of points, an advance of an animation on a computer screen, or a modification of a sound. When the EEG is related to symptoms of a disease, it has been shown that neurofeedback techniques can have a therapeutic effect, as is the case with attention

deficit disorder with hyperactivity (Micoulaud-Franchi et al. 2011) or epilepsy (Micoulaud-Franchi et al. 2014).

4.6.2 Implicit BCI

In contrast with explicit BCI, the implicit BCI is not based on the principle of operant conditioning. The feedback in implicit BCI is used to optimize the interaction with an external device by directly modulating the brain activity and the cognitive activity of the subject (George and Lécuyer 2010). Implicit BCI does not require a learning period. In practice, the subject does not have to try to control intentionally his EEG. The EEG signal is recorded, processed in real time to extract the information of interest (e.g. power spectral EEG or ERP) corresponding to the subject's cognitive activity, and transmitted to the external electronic device to modulate and optimize the interaction between the device and the user.

This configuration fits with some non-traditional neurofeedback therapeutics that do not require specific cognitive tasks and are supposed to directly modulate the brain activity of the subject in order to optimize brain dynamics, although this remains largely hypothetical. Thus, unlike traditional neurofeedback approaches presented in the previous section, these non-traditional neurofeedback approaches have a very low level of therapeutic and clinical evidence (Micoulaud-Franchi et al. 2013).

4.6.3 Towards an Intuitive Control Using Semiotic-Based BCI

From the two approaches inspired by previous theoretical frameworks from cognitive neuroscience (Sect. 4.2), we propose a prospective view on a sound synthesis control strategy based on BCI. We reflect on whether EEG BCI would be helpful to increase the intuitiveness of control with the sound synthesizer. For a didactic perspective, we suggest to describe explicit and implicit BCI, respectively, from the representational-computational and from the enactive points of view.

We stress that in the explicit BCI, the user controls the external electronic device (positive reinforcement) as if it was an external object. In some way, there is a gap between the information of interest extracted from the recorded EEG activity and the positive reinforcement. The information feedback could be given to the subject by any kind of signal. The positive reinforcement mainly is useful for the learning process and for determining a success rate and is close to an error function (Sect. 4.4.1). We think that in many cases, explicit BCI does not permit to create recurrent sensorimotor patterns (from the enactive point of view) that enable action to be guided by the direct perception of the stimulus, which could be a limitation in the intuitiveness of BCI controllability.

We stress that in the Implicit BCI, the user and his/her brain is involved in an enactive process. In some way, there is a direct link between the information of interest extracted from the recorded EEG and the feedback. This feedback is not a

positive reinforcement as defined by the operant-conditioning model. In fact, the aim of the feedback is not to inform the subject about the cognitive strategies that he/she develops during the learning process, but to directly influence the brain activity (and thus the EEG). Any kind of feedback cannot be used, but only those with the desired effect on the brain and the cognitive activity in order to enhance the interaction and the intuitiveness of the system.

Therefore, in the context of sound synthesis, a control strategy involving the use of explicit or implicit BCI would necessitate different mapping strategies. From a conceptual point of view, we stress that explicit and implicit BCI involve different levels of semiotic relation, i.e., the relation between the feedback and the meaning that the subject attributes to a sound. These two scenarios are discussed in the following paragraphs.

In the case of explicit BCI as defined above, the subject would have to control his/her cognitive activity to change his/her EEG and thus to control a specific parameter of the sound synthesizer. No semiotic relation between the EEG, the effect of the synthesized sound on the EEG, and the sound perception is therefore needed. In other words, the subject has to do something that is not necessarily related to the semiotics of the perceived synthesized sound to control the synthesizer. More so, an external algorithm is used to interpret the information of interest extracted from the EEG and to control the electronic device. For example, paying attention to a target to produce a P300 component that will be processed by the BCI and arbitrarily associated with a control parameter according to the output of the algorithm and to a success rate (Fig. 4.2). This situation that necessitates a certain expertise acquired during a learning period seems to be quite close to sound synthesis based on the physical or signal modelling of sound vibrations (Sect. 4.3).

In the case of implicit BCI as defined above, the aim would be to enhance the quality and the intuitiveness of the sound synthesizer by taking into account the EEG induced by the sound. Thus, a strict semiotic relation between the EEG and the influence of sounds on the EEG should be known. In other words, we need to understand the neural bases of sound semiotics ("electrophysiological data" in Fig. 4.2) to implement this information in an implicit BCI process dedicated to the sound synthesizer. We propose to call it "semiotic-based BCI". In this context, the results obtained from previous EEG experiments presented in Sect. 4.5 constitute an interesting starting point for the design of such a mapping strategy. This approach seems to be quite close to sound synthesis based on the modelling of perceptual effects, which does not necessitate a learning period (Sect. 4.3). This intuitive control implies that perceptual and cognitive aspects are taken into account in order to understand how a sound is perceived and interpreted. As shown in Fig. 4.2, a loop is thus designed between perception and action through the intuitive control of the sound synthesizer (Sect. 4.2). Implicit BCI offers the possibility of a second loop, between the sound effect on the EEG and the sound synthesizer that is likely to optimize the sound effect on both the perceptual judgment and the Implicit BCI.

4.7 Conclusion

To date, the design of a control strategy of sound synthesis processes that uses a BCI is still a challenging perspective. As discussed in (Väljamäe et al. 2013), a synthesis control of sounds directly from the brain through the measurement of its cerebral activity is still in its early stages. In particular, the mapping between electrophysiological signal features and synthesis parameters is generally validated on the basis on different metrics depending on applications. However, the definition of such metrics implies a given conception on the way we interact with the surrounding world.

To broach this issue, we introduced two conceptual approaches inspired from the representational-computational and the enactive paradigms from cognitive neuroscience. In light of these paradigms, we revisited the existing main approaches for synthesis and control of sounds. In fact, the viewpoints adopted to synthesize sounds are intricately underpinned by paradigms that differ in the epistemological positions of the observer (from a third or a first-person position) and have a substantial consequence on the design of a control strategy (cf. Figure 4.2). On one hand, synthesis processes based on the modelling of physical sources (from either the mechanical behaviour or the resulting vibration) are controlled by physical or signal parameters. This approach is based on the existence of a correct representation of the physical world and introduces the notion of an error function between the model and the physical reality as a quality criterion. Therefore, it requires a certain expertise from the end-user. On the other hand, synthesis processes based on the modelling of perceptual effects involve the identification of invariant sound morphologies specific to given perceptual attributes of the sound source. This approach assumes the emergence of an embodied auditory world from an enactive process. The perceptual judgments are considered as a quality criterion for the model, leading to the design of a more intuitive control.

By associating these conceptual and pragmatic considerations, we proposed a prospective view on the methodology to be used to design a BCI control. For the sake of illustration, we treated limited aspects of BCIs by addressing explicit BCI from the representational-computational point of view and implicit BCI from the enactive point of view. Actually, we are aware that the frontier between explicit and implicit BCI might be difficult to establish and less didactic than what this article presents. Indeed, the implicit communication channel might sometimes be used in an explicit way (George and Lécuyer 2010), and inversely brain plasticity can enable the participant to make use of the training experienced from the explicit BCI to generate implicit recurrent sensorimotor patterns (Bach-y-Rita and Kercel 2003). With current apparatus performances, the rate of transfer information between the BCI and the device is quite limited and the final task has to be defined accordingly. While this technique may represent a restricted interest for healthy users (in some cases, it would be easier to directly control the device manually), it constitutes a relevant medium for medical applications and can be used as a substitutional device for diseases. In the implicit BCI, the control is included in an optimization system in

which the electrophysiological data supplies further information about the way the user perceives the sound (beyond verbal labels or gestures for instance). In contrast with the explicit BCI, this configuration is well adapted to intuitive synthesis control. Therefore, we suggested a "semiotic-based BCI" founded on identified links between the brain activity and invariant signal morphologies reflecting the attribution of sense to a sound that may enhance the interactivity and the intuitiveness of the system.

4.8 Questions

1. What are the characteristics of the representational-computational paradigm of perception?
2. What are the characteristics of the enactive paradigm of perception?
3. What is the difference between physical and signal sound synthesis models?
4. What are the main limitations of the use of physical models for sound synthesis?
5. How can the invariant sound morphologies be determined?
6. Which invariant sound morphologies are related to the perception of material in an impact sound?
7. Which aspects should be taken into account in the design of a control strategy based on a representational-computational or an enactive paradigm?
8. What are the characteristics of explicit (or active) BCI?
9. What are the characteristics of implicit (or passive) BCI?
10. What is the purpose of offering intuitive control of sound synthesis processes using BCI?

References

Aramaki M, Besson M, Kronland-Martinet R, Ystad S (2009) Timbre perception of sounds from impacted materials: behavioral, electrophysiological and acoustic approaches. In: Ystad S, Kronland-Martinet R, Jensen K (eds) Computer music modeling and retrieval—genesis of meaning of sound and music, vol 5493., LNCSSpringer, Berlin, Heidelberg, pp 1–17

Aramaki M, Besson M, Kronland-Martinet R, Ystad S (2011) Controlling the perceived material in an impact sound synthesizer. IEEE Trans Audio Speech Lang Process 19(2):301–314

Aramaki M, Gondre C, Kronland-Martinet R, Voinier T, Ystad S (2010a) Imagine the sounds: an intuitive control of an impact sound synthesizer. In: Ystad S, Aramaki M, Kronland-Martinet R, Jensen K (eds) Auditory display, vol 5954., Lecture notes in computer scienceSpringer, Berlin, Heidelberg, pp 408–421

Aramaki M, Marie C, Kronland-Martinet R, Ystad S, Besson M (2010b) Sound categorization and conceptual priming for nonlinguistic and linguistic sounds. J Cogn Neurosci 22(11):2555–2569

Arfib D (1979) Digital synthesis of complex spectra by means of multiplication of non-linear distorted sine waves. J Audio Eng Soc 27:757–768

Atal BS, Hanauer SL (1971) Speech analysis and synthesis by linear prediction of the speech wave. J Acoust Soc Am 50(2B):637–655

Avanzini F, Serafin S, Rocchesso D (2005) Interactive simulation of rigid body interaction with friction-induced sound generation. IEEE Trans Speech Audio Process 13(5):1073–1081

Bach-y-Rita P, Kercel W (2003) Sensory substitution and the human-machine interface. Trends in Cogn Sci 7:541–546

Ballas JA (1993) Common factors in the identification of an assortment of brief everyday sounds. J Exp Psychol Hum Percept Perform 19(2):250–267

Bensa J, Jensen K, Kronland-Martinet R (2004) A hybrid resynthesis model for hammer-strings interaction of piano tones. EURASIP J Appl Sig Process 7:1021–1035

Bilbao S (2009) Numerical sound synthesis: finite difference schemes and simulation in musical acoustics. Wiley, Chichester, UK

Castle PC, van Toller S, Milligan G (2000) The effect of odour priming on cortical EEG and visual ERP responses. Int J Psychophysiol 36:123–131

Chaigne A (1995) Trends and challenges in physical modeling of musical instruments, In: Proceedings of the international congress on acoustics', Trondheim, Norway

Chowning J (1973) The synthesis of complex audio spectra by means of frequency modulation. J Audio Eng Soc 21:526–534

Conan S, Aramaki M, Kronland-Martinet R, Thoret E, Ystad S (2012) Perceptual differences between sounds produced by different continuous interactions. Proceedings of the 11th Congrès Français d'Acoustique. Nantes, France, pp 409–414

Conan S, Aramaki M, Kronland-Martinet R, Ystad S (2013) Post-proceedings 9th International Symposium on Computer Music Modeling and Retrieval (CMMR 2012). Lecture notes in computer science, vol 7900. Springer, Berlin, Heidelberg, chapter Intuitive Control of Rolling Sound Synthesis

Conan S, Thoret E, Aramaki M, Derrien O, Gondre C, Kronland-Martinet R, Ystad S (2013) Navigating in a space of synthesized interaction-sounds: rubbing, scratching and rolling sounds. In: Proceedings of the 16th international conference on digital audio effects (DAFx-13), Maynooth, Ireland

Cook PR (1992) A meta-wind-instrument physical model, and a meta-controller for real-time performance control. In: Proceedings of the international computer music conference, pp 273–276

Daltrozzo J, Schön D (2009) Conceptual processing in music as revealed by N400 effects on words and musical targets. J Cogn Neurosci 21:1882–1892

de Saussure F (1955) Cours de linguistique générale. Payot, Paris

Flanagan JL, Coker CH, Rabiner LR, Schafer RW, Umeda N (1970) Synthetic voices for computer. IEEE Spectr 7:22–45

Gaver WW (1993a) How do we hear in the world? Explorations of ecological acoustics. Ecol Psychol 5(4):285–313

Gaver WW (1993b) What in the world do we hear? An ecological approach to auditory source perception. Ecol Psychol 5(1):1–29

George L, Lécuyer A (2010) An overview of research on "passive" brain-computer interfaces for implicit human-computer interaction. In: International conference on applied bionics and biomechanics ICABB 2010—workshop W1 Brain-Computer Interfacing and Virtual Reality, Venezia, Italy

Gibson JJ (1986) The ecological approach to visual perception, Lawrence Erlbaum Associates

Giordano BL, McAdams S (2006) Material identification of real impact sounds: effects of size variation in steel, wood, and plexiglass plates. J Acoust Soc Am 119(2):1171–1181

Gygi B, Kidd GR, Watson CS (2007) Similarity and categorization of environmental sounds. Percept Psychophys 69(6):839–855

Gygi B, Shafiro V (2007) General functions and specific applications of environmental sound research. Front Biosci 12:3152–3166

Hermes DJ (1998) Synthesis of the sounds produced by rolling balls. Internal IPO report no. 1226, IPO, Center for user-system interaction, Eindhoven, The Netherlands

Holcomb PJ, McPherson WB (1994) Event-related brain potentials reflect semantic priming in an object decision task. Brain and Cogn 24:259–276. http://forumnet.ircam.fr/product/modalys/?lang=en (n.d.). http://www-acroe.imag.fr/produits/logiciel/cordis/cordis_en.html (n.d.)
Husserl E (1950) Idées directrices pour une phénoménologie, Gallimard. J New Music Res, special issue "enaction and music" (2009), 38(3), Taylor and Francis, UK
Karjalainen M, Laine UK, Laakso T, Vilimtiki V (1991) Transmission-line modeling and real-time synthesis of string and wind instruments. In: I. C. M. Association (ed) Proceedings of the international computer music conference, Montreal, Canada, pp 293–296
Koelsch S, Kasper E, Sammler D, Schulze K, Gunter T, Friederici A (2004) Music, language and meaning: brain signatures of semantic processing. Nat Neurosci 7(3):302–307
Kronland-Martinet R (1989) Digital subtractive synthesis of signals based on the analysis of natural sounds. A.R.C.A.M. (ed), Aix en Provence
Kronland-Martinet R, Guillemain P, Ystad S (1997) Modelling of natural sounds by time-frequency and wavelet representations. Organ Sound 2(3):179–191
Kutas M, Hillyard SA (1980) Reading senseless sentences: brain potentials reflect semantic incongruity. Science 207:203–204
Lachaux JP, Rodriguez E, Martinerie J, Varela F (1999) Measuring phase synchrony in brain signals. Hum Brain Mapp 8:194–208
Lalande A (1926) Vocabulaire technique et critique de la philosophie, Edition actuelle, PUF n quadrige z 2002
Le Brun M (1979) Digital waveshaping synthesis. J Audio Eng Soc 27:250–266
Makhoul J (1975) Linear prediction, a tutorial review. In: Proceedings of the IEEE, vol 63. pp 561–580
Matyja JR, Schiavio A (2013) Enactive music cognition: background and research themes. Constr Found 8(3):351–357. http://www.univie.ac.at/constructivism/journal/8/3/351.matyja
McAdams S (1999) Perspectives on the contribution of timbre to musical structure. Comput Music J 23(3):85–102
McAdams S, Bigand E (1993) Thinking in sound: the cognitive psychology of human audition, Oxford University Press, Oxford
Merer A, Ystad S, Kronland-Martinet R, Aramaki M (2011) Abstract sounds and their applications in audio and perception research. In: Ystad S, Aramaki M, Kronland-Martinet R, Jensen K (eds) Exploring music contents, vol 6684., Lecture notes in computer scienceSpringer, Berlin, Heidelberg, pp 176–187
Micoulaud-Franchi JA, Bat-Pitault F, Cermolacce M, Vion-Dury J (2011) Neurofeedback dans le trouble déficit de l'attention avec hyperactivité : de l'efficacité à la spécificité de l'effet neurophysiologique. Annales Médico-Psychologiques 169(3):200–208
Micoulaud-Franchi JA, Cermolacce M, Vion-Dury J, Naudin J (2013) Analyse critique et épistémologique du neurofeedback comme dispositif thérapeutique. le cas emblématique du trouble déficit de l'attention avec hyperactivité', L'évolution psychiatrique
Micoulaud-Franchi J, Lanteaume L, Pallanca O, Vion-Dury J, Bartolomei F (2014) Biofeedback et épilepsie pharmacorésistante : le retour d'une thérapeutique ancienne ? Revue Neurologique 170(3):187–196
Nadeau R (1999) Vocabulaire technique et analytique de l'épistémologie, PUF
Nijholt A (2009) BCI for games: a 'state of the art' survey. In: Stevens SM, Saldamarco SJ (eds) LNCS, vol 5309. Springer, Berlin, pp 225–228
O'Brien JF, Shen C, Gatchalian CM (2002) Synthesizing sounds from rigid-body simulations. In: Press A (ed) The ACM SIGGRAPH 2002 symposium on computer animation, pp 175–181
Orgs G, Lange K, Dombrowski J, Heil M (2006) Conceptual priming for environmental sounds and words: An ERP study. Brain Cogn 62(3):267–272
Pai DK, van den Doel K, James DL, Lang J, Lloyd JE, Richmond JL, Yau SM (2001) Scanning physical interaction behavior of 3D objects. In: Proceedings of SIGGRAPH 2001, computer graphics proceedings, annual conference series, pp 87–96

Petitmengin C, Bitbol M, Nissou JM, Pachoud B, Curalucci H, Cermolacce M, Vion-Dury J (2009) Listening from within. J Conscious Stud 16:252–284

Rabiner LR, Gold B (1975) Theory and application of digital signal processing. Prentice Hall, Englewood Cliffs, NJ

Rath M, Rocchesso D (2004) Informative sonic feedback for continuous human–machine interaction—controlling a sound model of a rolling ball. IEEE Multimedia Spec Interact Sonification 12(2):60–69

Risset JC (1965) Computer study of trumpet tones. J Acoust Soc Am 33:912

Roads C (1978) Automated granular synthesis of sound. Comput Music J 2(2):61–62

Rugg MD, Coles MGH (1995) Electrophysiology of mind. Event-related brain potentials and cognition, number 25. In: 'Oxford Psychology', Oxford University Press, chapter The ERP and Cognitive Psychology: Conceptual issues, pp 27–39

Schaeffer P (1966) Traité des objets musicaux, du Seuil (ed)

Schön D, Ystad S, Kronland-Martinet R, Besson M (2010) The evocative power of sounds: conceptual priming between words and nonverbal sounds. J Cogn Neurosci 22(5):1026–1035

Smith JO (1992) Physical modeling using digital waveguides. Comput Music J 16(4):74–87

Stoelinga C, Chaigne A (2007) Time-domain modeling and simulation of rolling objects. Acta Acustica united Acustica 93(2):290–304

Thoret E, Aramaki M, Gondre C, Kronland-Martinet R, Ystad S (2013) Controlling a non linear friction model for evocative sound synthesis applications. In: Proceedings of the 16th international conference on digital audio effects (DAFx-13), Maynooth, Ireland

Thoret E, Aramaki M, Kronland-Martinet R, Velay J, Ystad S (2014) From sound to shape: auditory perception of drawing movements. J Exp Psychol Hum Percept Perform

Thoret E, Aramaki M, Kronland-Martinet R, Ystad S (2013) Post-proceedings 9th International Symposium on Computer Music Modeling and Retrieval (CMMR 2012), number 7900. In: Lecture notes in computer science, Springer, Berlin, Heidelberg, chapter reenacting sensori-motor features of drawing movements from friction sounds

Väljamäe A, Steffert T, Holland S, Marimon X, Benitez R, Mealla S, Oliveira A, Jordà S (2013) A review of real-time EEG sonification research. In: Proceedings of the 19th international conference on auditory display (ICAD 2013), Lodz, Poland, pp 85–93

van den Doel K, Kry PG, Pai DK (2001) Foleyautomatic: physically-based sound effects for interactive simulation and animation. In: Proceedings of SIGGRAPH 2001, computer graphics proceedings, annual conference series, pp 537–544

Van Petten C, Rheinfelder H (1995) Conceptual relationships between spoken words and environmental sounds: event-related brain potential measures. Neuropsychologia 33 (4):485–508

Vanderveer NJ (1979) Ecological acoustics: human perception of environmental sounds, PhD thesis, Georgia Inst. Technol

Varela F (1989) Invitation aux sciences cognitives. Seuil, Paris

Varela F (1996) Neurophenomenology: a methodological remedy for the hard problem. J Conscious Stud 3:330–335

Varela F, Thompson E, Rosch E (1991) The embodied mind: cognitive science and human experience. MIT Press, Cambridge, MA, USA

Verron C, Aramaki M, Kronland-Martinet R, Pallone G (2010) A 3D immersive synthesizer for environmental sounds. IEEE Trans Audio Speech Lang Process 18(6):1550–1561

Verron C, Pallone G, Aramaki M, Kronland-Martinet R (2009) Controlling a spatialized environmental sound synthesizer. Proceedings of the IEEE workshop on applications of signal processing to audio and acoustics (WASPAA). New Paltz, NY, pp 321–324

Viviani P (2002) Motor competence in the perception of dynamic events: a tutorial. In: Prinz W, Hommel B (eds) Common mechanisms in perception and action. Oxford University Press, New York, NY, pp 406–442

Viviani P, Redolfi M, Baud-Bovy G (1997) Perceiving and tracking kinaesthetic stimuli: further evidence for motor-perceptual interactions. J Exp Psychol Hum Percept Perform 23:1232–1252

Viviani P, Stucchi N (1992) Biological movements look uniform: evidence of motor-perceptual interactions. J Exp Psychol Hum Percept Perform 18:603–623
Warren WH, Verbrugge RR (1984) Auditory perception of breaking and bouncing events: a case study in ecological acoustics. J Exp Psychol Hum Percept Perform 10(5):704–712
Wolpaw JR, Birbaumer N, McFarland DJ, Pfurtscheller G, Vaughan TM (2002) Brain-computer interfaces for communication and control. Clin Neuro physiol 113:767–791
Ystad S, Voinier T (2001) A virtually-real flute. Comput Music J 25(2):13–24