HAL open science

Vers une approche acoustique et cognitive de la sémiotique des objets sonores

Sølvi Ystad, Richard Kronland-Martinet, Danièle Schön, Mireille R Besson

To cite this version:

Sølvi Ystad, Richard Kronland-Martinet, Danièle Schön, Mireille R Besson. Vers une approche acoustique et cognitive de la sémiotique des objets sonores. Journées USTS, 2005, Nantes, France. hal-00093346

HAL Id: hal-00093346

https://hal.science/hal-00093346v1

Submitted on 24 Aug 2023

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.

Vers une approche acoustique et cognitive de la sémiotique des objets sonores

Sølvi Ystad[1], Richard Kronland-Martinet[1],
Daniele Schön[2] et Mireille Besson[2]

[1]Equipe Modélisation, Synthèse et Contrôle des Signaux Sonores et Musicaux,
CNRS - Laboratoire de Mécanique et d'Acoustique, Marseille
[2]Equipe Langage et Musique,
CNRS - Institut de Neurosciences Cognitives de la Méditerranée, Marseille

Introduction

Ce travail a pour but de mieux comprendre les mécanismes perceptifs et cognitifs mis en œuvre lors de l'écoute des sons, et en particulier, d'étudier les relations entre les traitements cognitifs associés à la sémantique dans le langage et à la sémiotique[1] des objets sonores (i.e. comment attribue-t-on un sens au son). Il vise également à mieux cerner les relations entre la structure acoustique des sons et l'information perceptive et cognitive qu'ils véhiculent. Pour ce faire nous proposons une approche pluridisciplinaire reposant sur la collaboration entre l'équipe « Modélisation, Synthèse et Contrôle des Signaux Sonores et Musicaux » du LMA-CNRS, (Ystad et Kronland-Martinet) et l'équipe « Langage, Musique et motricité » de l'INCM-CNRS (Schön et Besson).

Si des grandes avancées méthodologiques permettent aujourd'hui de construire par synthèse numérique des sons complexes, les indices perceptifs et cognitifs des sons sont encore relativement peu connus. Ainsi peut-on par exemple re-synthétiser le son produit par différents pianos (Bensa, 2003) sans pour autant bien comprendre ce qui fait la différence auditive entre un bon et un excellent piano (Bensa *et al.*, 2004). De même est-il extrêmement difficile de contrôler un système de synthèse sonore à partir d'une description sémantique des sources (par exemple : bruit d'impact d'un marteau métallique sur une plaque de verre de petite dimension) (Gaver, W. 1993, Aramaki *et al.*, 2004). Ces exemples illustrent le besoin de mieux comprendre les relations entre la structure des sons et leur perception. D'un autre coté, la communauté des Neurosciences Cognitives manifeste un vif intérêt pour comprendre les relations entre les sons (notamment ceux de la parole et la musique) et les processus cérébraux mis en jeu lors de l'écoute (Patel *et al.*, 1998 ; Overy, 2003 ; Schön *et al.*, 2004). Ces travaux, dont le but principal est de mieux définir l'architecture fonctionnelle du cerveau, apportent une nouvelle vision des

[1]i.e. étude des pratiques signifiantes dans les divers domaines de la communication (Petit Larousse).

phénomènes liés à l'écoute des sons (Griffiths *et al.*, 2004) tout en proposant un ensemble de méthodes permettant des mesures objectives de l'activité cérébrale (potentiels évoqués, IRMf, etc.) (Rugg *et al.*, 1995 ; Buxton, 2002). Dans l'étude présentée ici, l'objectif est de mieux comprendre comment le cerveau attribue un sens aux sons et si ces processus sont similaires pour les sons et les mots. On se basera pour cela sur des études antérieures liées à la sémantique du langage (Kutas et Hillyard, 1980), en comparant l'activité cérébrale évoquée par des mots et par des sons.

Afin de dissocier les sons des associations éventuelles liées à notre culture, nous avons construit un corpus sonore sans relation directe avec des sons habituels (musicaux ou de l'environnement) en effectuant un échantillonnage de l'espace sonore basé sur la classification proposée par Pierre Schaeffer. Un corpus de 45 objets sonores issus pour l'essentiel de matériaux « concrets » et représentatifs des classes de sons proposées par Schaeffer a ainsi été constitué. Dans la première partie de l'expérience, les sons ont été présentés à 7 sujets qui devaient associer un (ou plusieurs) mot(s) à chaque son. Les mots les plus fréquemment évoqués ont ensuite été sélectionnés, formant ainsi 45 paires mots/sons associés. Une deuxième liste de mots, dont la signification n'est pas associée aux sons a ensuite été constituée étendant ainsi le corpus à 45 triplets "mots/sons associés/sons non associés". Dans la seconde partie de l'expérience, les mots ont été présentés visuellement aux sujets de façon aléatoire, suivis par la présentation auditive du son associé ou non au mot visualisé. La tâche du sujet consistait à décider, le plus rapidement possible, si la paire mot/son était compatible ou non. Les Temps de Réaction (TRs) ainsi que les potentiels électriques évoqués (PEs), mesurés à l'aide d'électrodes placées sur la tête du sujet, ont été recueillis au cours de l'expérience. Nous présentons ci-dessous les premiers résultats de cette expérience, ainsi que les perspectives qui en découlent.

Méthodologie

Plusieurs approches sont envisageables pour aborder le problème de la sémiotique des objets sonores. Dans cette étude, nous avons choisi d'associer à des analyses acoustiques des mesures électrophysiologiques, de façon à bénéficier de mesures objectives de l'activité cérébrale pour lesquelles de nombreux travaux, notamment dans le cas du langage, pourront servir de support aux interprétations.

Méthode ERP

En plaçant des électrodes sur le scalp de volontaires humains, il est possible d'enregistrer les variations de l'activité cérébrale, connues sous le nom d'électroencephalogramme (EEG) (Berger, 1929). L'EEG ayant une faible amplitude (de l'ordre de 100 microvolts), les signaux analogiques sont amplifiés, puis convertis en signaux numériques (Figure 1).

N'importe quel stimulus, tel qu'un son, une lumière, une odeur, etc., provoque des changements ou variations du potentiel dans l'EEG. Néanmoins, ces variations sont très petites (de l'ordre de 10 μV), par rapport à l'activité électrique de base. De façon à extraire le signal utile, il est alors nécessaire de synchroniser l'enregistrement EEG par rapport au début de la stimulation, et de moyenner un grand nombre d'essais obtenus dans les mêmes conditions expérimentales et pour des stimuli de même type. Ainsi, le signal – c'est-à-dire les variations de potentiel évoquées par la stimulation (Potentiels Évoqués ou PEs) – pourront émerger de l'activité électrique de base (Figure 2).

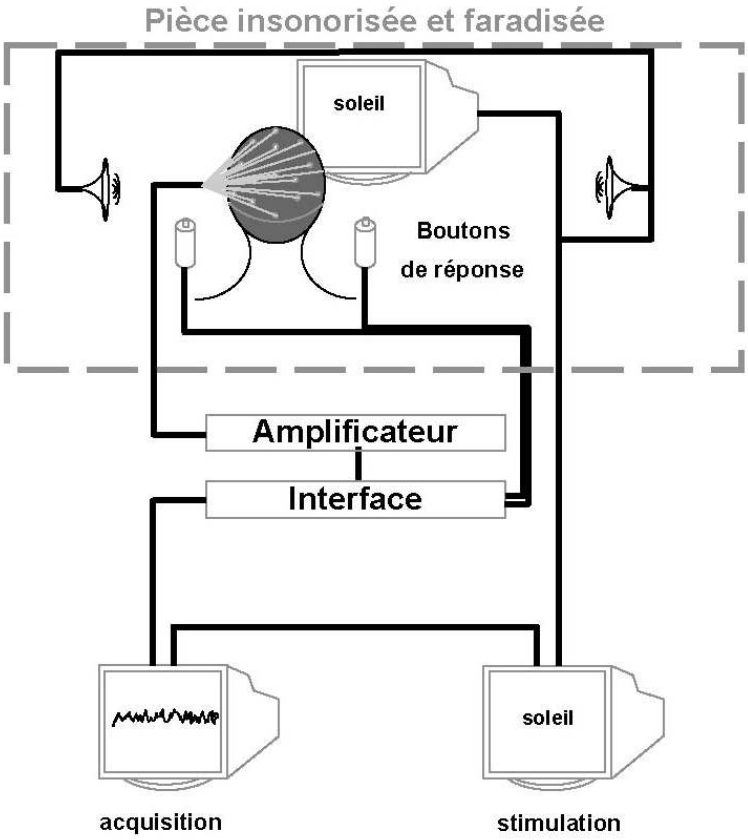

Figure 1. Système d'acquisition de données issues de mesures électrophysiologiques

Typiquement, dans des expériences perceptives et cognitives, 20 à 30 stimuli pour chaque condition expérimentale sont nécessaires pour obtenir un rapport signal sur bruit satisfaisant par « moyennage ». Les Potentiels Évoqués contiennent une série de déflections positives et négatives par rapport à la ligne de base, appelés composantes. La ligne de base correspond au niveau moyen de l'activité cérébrale pendant 100 ou 200 millisecondes (ms) avant la stimulation. Les composantes sont définies par leur polarité (négative, N, ou positive, P), leur latence à partir du début du stimulus (100, 200, 300, 400 ms, etc.), leur distribution sur le scalp (localisation de l'amplitude maximale sur le scalp), ainsi que leur sensibilité liée aux facteurs expérimentaux. Typiquement, les PEs enregistrés dans des expériences perceptives et cognitives comprennent des composantes précoces qui reflètent les étapes sensorielles et perceptives du traitement de l'information (e.g., P100, N200, etc.) et des composantes plus tardives qui reflètent les étapes cognitives et décisionnelles (P300, N400, etc.). Les PEs sont analysés dans des fenêtres de latence différentes, centrées autour de l'amplitude maximale de la composante étudiée. Par exemple, pour analyser la composante N400, une fenêtre de latence comprise entre 300 et 600 ms est généralement utilisée. L'amplitude de la composante est obtenue à partir de la moyenne des acquisitions dans ce rang de latence.

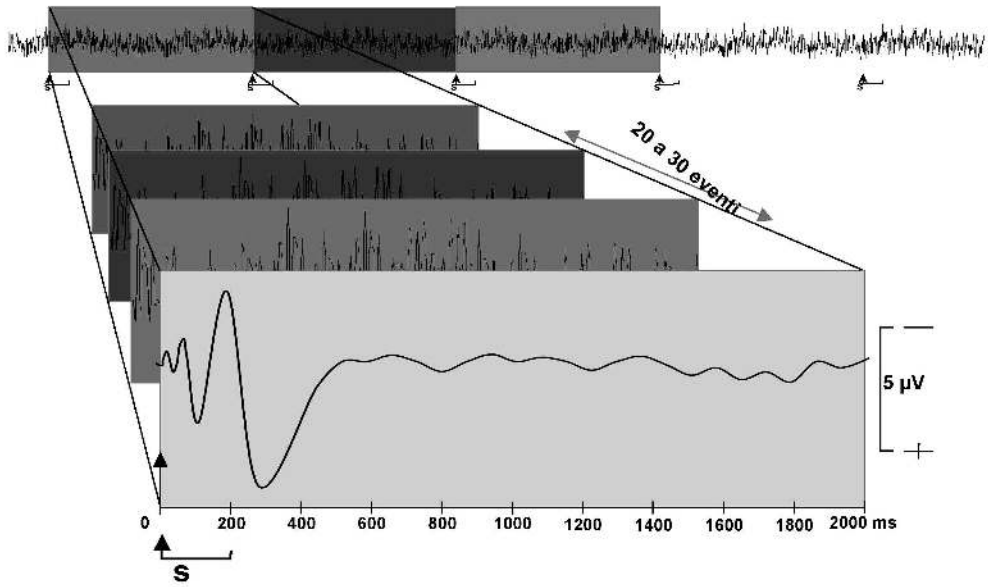

Figure 2. Moyennage de données électrophysiologiques

Traitement sémantique dans le langage

La méthode des Potentiels Évoqués a été largement utilisée dans de nombreuses études portant sur la sémantique du langage. Cette approche a permis en 1980 à Kutas et Hillyard (M. Kutas *et al.*, 1980) de démontrer l'existence d'une déflection négative de l'activité cérébrale électrique (composante N400), dont l'amplitude est modulée par le degré de congruence sémantique. Ce résultat est illustré sur la Figure 3. Dans cet exemple, des phrases dont le dernier mot peut être congruent ou incongru, ont été présentées à des sujets pendant que leur activité cérébrale était mesurée. Les résultats ont montré une augmentation de l'amplitude de la composante N400 lorsque le dernier mot était incongru. De plus, cette augmentation dépend du degré d'incongruité. Ainsi, le mot terminal « narines » dans la phrase « Il porte sa fille dans ses narines » provoque une forte augmentation de l'amplitude de la composante N400, alors que le mot terminal « mains » présente une plus petite augmentation. Par contre, quand le mot terminal « bras » est utilisé, l'amplitude de la composante N400 est très faible. Cet exemple montre que la méthode des Potentiels Évoqués permet de mesurer le degré de congruence sémantique d'un mot dans un contexte.

D'autres études sur des couples de mots isolés appartenant ou non à la même catégorie sémantique ont également mis en évidence l'existence de la composante N400. Ainsi, lorsque deux mots appartiennent à la même catégorie sémantique (par exemple « chien-chat » ; « docteur-infirmière »), le second mot de la paire évoque une composante N400 de plus faible amplitude que s'il appartient à une catégorie sémantique différente du premier mot (par exemple « poire-chat » ; « abeille-infirmière ») (Bentin *et al.*, 1985).

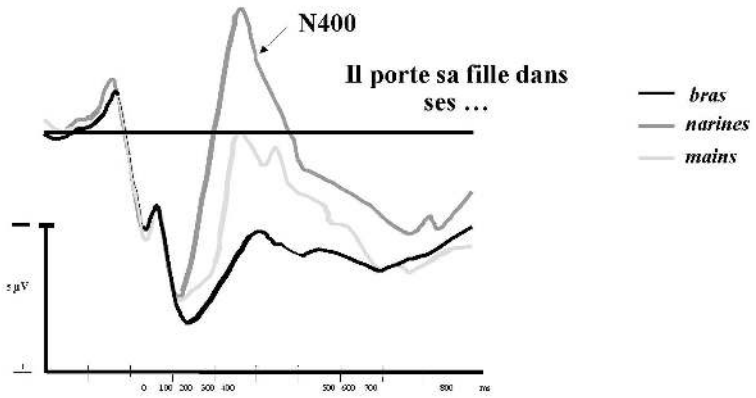

Figure 3. Mesure de congruence sémantique dans le langage

Traitement sémiotique des sons

Si de nombreuses études ont permis de mettre en évidence une activité cérébrale spécifique au traitement sémantique dans le langage, il n'en est pas de même pour la sémiotique des sons. Plusieurs études ont tenté de mettre en évidence des situations où la sémantique (ou sémiotique) sonore ferait l'objet du même traitement cognitif que la sémantique dans le langage. Une des difficultés de cette recherche réside dans le choix des stimuli. Dans une étude récente, Koelsch et ses collègues (Koelsch *et al.*, 2004) ont montré qu'en présentant des extraits de musique classique suivis par des mots judicieusement choisis, l'amplitude de la N400 liée au mot varie en fonction du type d'extrait musical qui le précède (Figure 4).

Bien que cette étude ait permis de montrer l'existence d'une composante N400 dans une expérience alliant musique et langage, ces résultats ne permettent pas d'affirmer que les sons (ou la musique) induisent le même type de traitement cognitif que la sémantique dans le langage. En effet, l'augmentation de l'amplitude N400 est provoquée par un mot non associé à un extrait musical et non par un son. De plus, l'utilisation d'extraits de musique classique peut refléter l'influence de notre culture sur ces résultats.

Vers une sémiotique de l'objet sonore

Contrairement à l'étude de Koelsch, dans l'étude présentée ici nous nous sommes focalisés sur l'étude de sons isolés, associés à des mots, de façon à analyser l'activité cérébrale en fonction de la cohérence entre le mot et le son.

De manière à dissocier les sons des associations éventuelles liées à notre culture, nous nous sommes efforcés de travailler avec des stimuli plus « neutres » qui devraient évoquer des réactions indépendamment de la culture des sujets. Ainsi, pour constituer un ensemble de sons variés et équilibrés, nous nous sommes basés sur une écoute acousmatique (sans référence à la source) et nous avons constitué un corpus sonore représentatif des sons « équilibrés » selon les critères de Pierre Schaeffer (Schaeffer, 1966). Les objets sonores ainsi construits diffèrent par leur « forme » (itératifs, continus, impulsifs, etc.) et leur

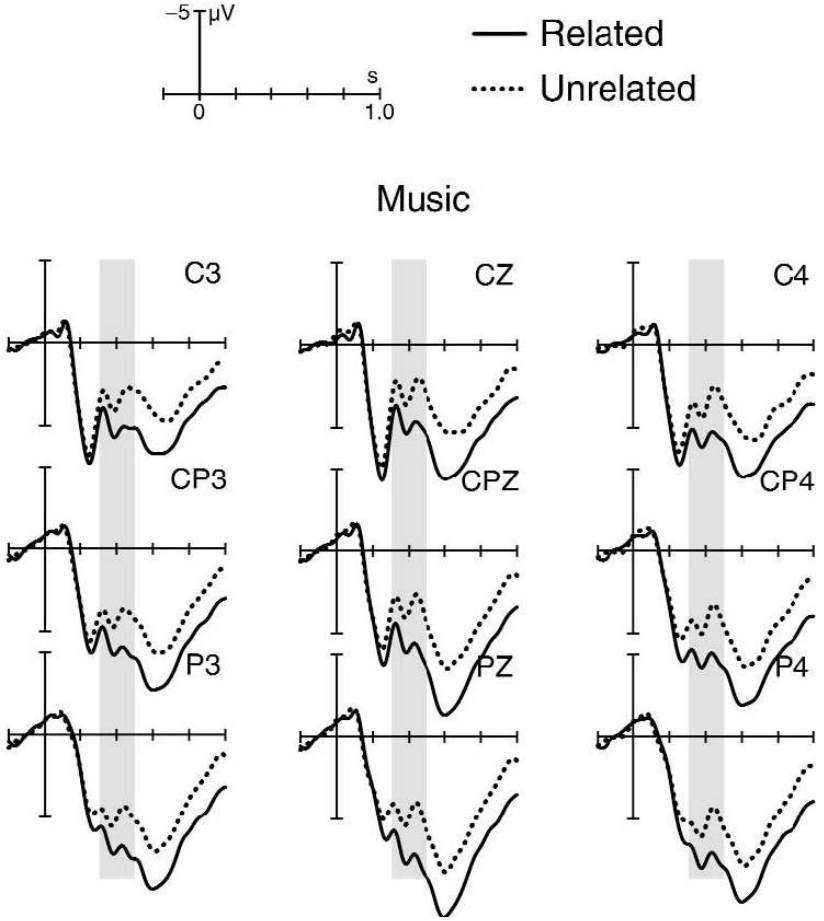

Figure 4. Potentiels Évoqués par un mot associé (tracé continu) ou non (tracé en pointillé) à l'extrait musical qui le précède (d'après Koelsch *et al.*, 2004)

« matière » (granuleux, inharmonique, etc.) (voir Tableau 1). 45 sons repartis dans 9 classes différentes de sons équilibrés ont ainsi été construits à partir de sons concrets et synthétiques.

	Sons tenus	Sons impulsifs	Sons itératifs
Hauteur définie	N	N'	N''
Hauteur complexe	X	X'	X''
Hauteur variée	Y	Y'	Y''

Tableau 1. Tableau des objets sonores « équilibrés » selon le classement de P. Schaeffer

Les sons ont été présentés à 11 sujets dont la tâche était d'associer un (ou plusieurs) mot(s) à chaque son. Les mots les plus fréquemment évoqués ont ensuite été choisis, formant ainsi 45 paires mots/sons associés. Une deuxième liste de mots, dont la signification n'est pas associée aux sons a ensuite été constituée formant ainsi un corpus de 45 triplets « mots/sons associés/sons non associés » (la liste complète des mots est donnée en annexe).

Déroulement de l'expérience électrophysiologique

Ainsi que nous l'avons décrit précédemment, l'activité cérébrale peut être mesurée en plaçant des électrodes sur le scalp des sujets volontaires. Dans notre expérience, un casque comprenant 32 électrodes a été utilisé. Le sujet regarde un écran d'ordinateur sur lequel un mot apparaît suivi après un intervalle de 500 ms par un son.

La tâche du sujet consiste à juger (le plus rapidement possible) si le son est relié au mot ou non. Il donne sa réponse en appuyant sur un des 2 boutons de réponse qu'il tient dans ses mains pendant l'expérience. Les données comportementales (pourcentage d'erreurs et Temps de Réaction, TR) ainsi que l'activité cérébrale (EEG) sont enregistrées. 45 paires mots/sons ont été présentées par condition.

Résultats

Les données comportementales montrent que le TR est relativement long et ne semble pas dépendre de la nature des couples (mots/sons associés et mots/sons non-associés). Le pourcentage d'erreur est dans l'ensemble faible et moins élevé pour les mots non reliés aux sons (5%) que pour les mots reliés aux sons (16%). Ainsi il serait plus facile de déterminer que les sons ne sont pas associés aux mots que de découvrir qu'ils le sont. Il est intéressant de noter que certains couples mots/sons sont jugés comme associés par tous les sujets alors que d'autres couples sont beaucoup plus ambigus (50% d'erreurs). Les résultats issus des mesures électrophysiologiques sont illustrés sur la Figure 5. Les 18 tracés sont obtenus à partir de moyennages (effectués sur 11 sujets) des acquisitions issues des conditions expérimentales mots/sons associés ou mots/sons non associés.

Les effets obtenus montrent une augmentation relative de négativité entre 300 et 600 ms lorsque le son n'est pas associé au mot qui le précède, par rapport à la condition où le son est associé au mot (voir agrandissement de l'effet observé sur une électrode centrale, Cz, sur la droite de la Figure 5). Ces résultats sont similaires à ceux obtenus avec des mots sémantiquement reliés et non reliés dans le langage.

Analyse acoustique

Dans le but de mieux comprendre le lien entre la structure des sons et leur perception, nous avons fait des analyses acoustiques approfondies des sons utilisés dans cette expérience. Les paramètres extraits des sons sont de deux types : des paramètres liés au signal (durée, distribution temps-fréquence de l'énergie, etc.) et des paramètres liés au timbre (temps d'attaque, brillance, etc.). Les Figures 6 et 7 représentent les spectrogrammes de 2 sons utilisés dans l'expérience. Le premier, associé à un son classé X selon Schaeffer (son entretenu et à hauteur complexe), est riche en composantes spectrales et varie peu au cours du temps. Cette caractérisation est générale pour les sons de la classe X

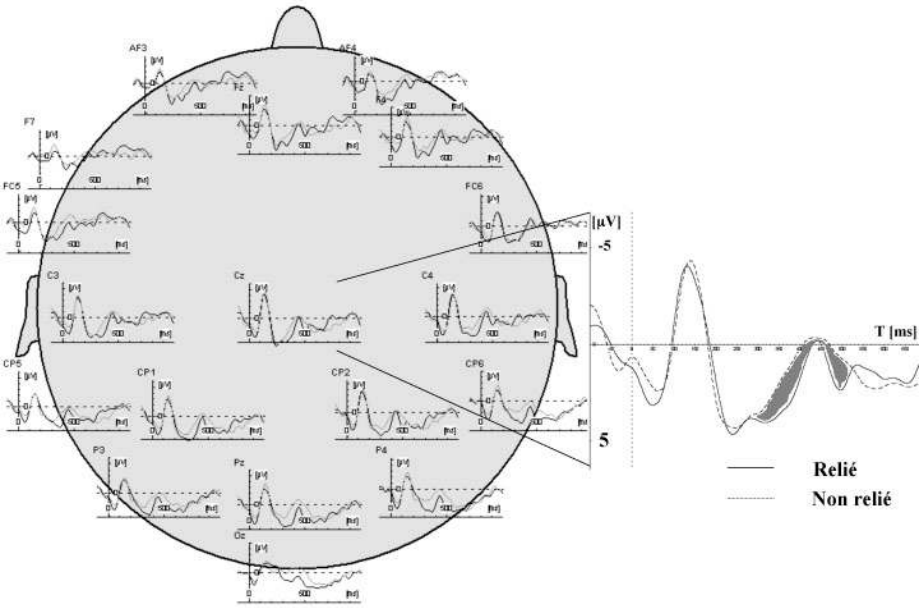

Figure 5. Potentiels évoques par le second stimulus (le son) et enregistrés à partir de différentes électrodes sur le scalp (la négativité est vers le haut)

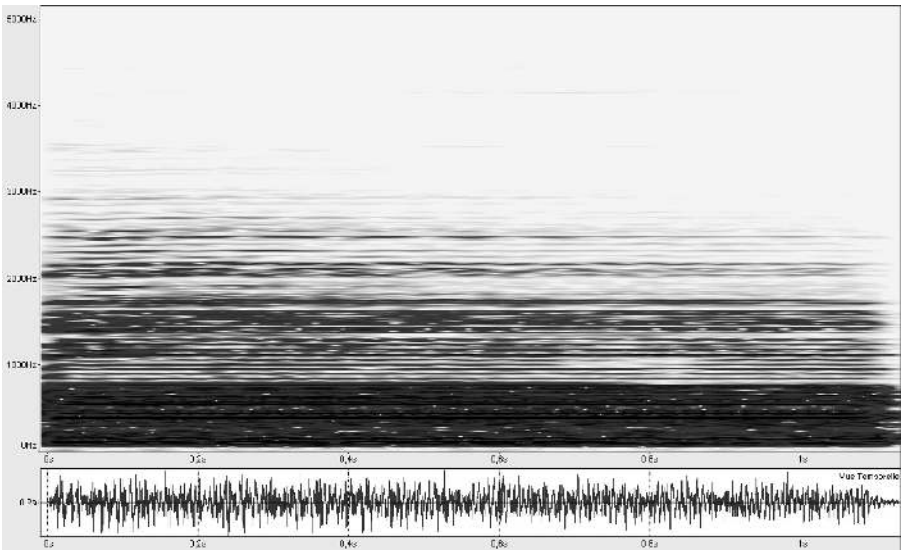

Figure 6. Spectrogramme d'un son entretenu à hauteur complexe (X selon la classification de Schaeffer)

qui sont très souvent associés à des mots décrivant des situations désagréables (frisson, froid, douleur, etc.). Le deuxième spectrogramme est associé à un son Y (son entretenu à hauteur variable). Il montre une plus faible densité spectrale et des composantes de haute fréquence qui s'amortissent rapidement. Ces composantes manifestent de plus une tendance non-stationnaire avec des évolutions de fréquence au cours du temps. Les sons de cette classe sont très souvent associés aux mouvements (rebond, monté, roulement, rotation, etc.).

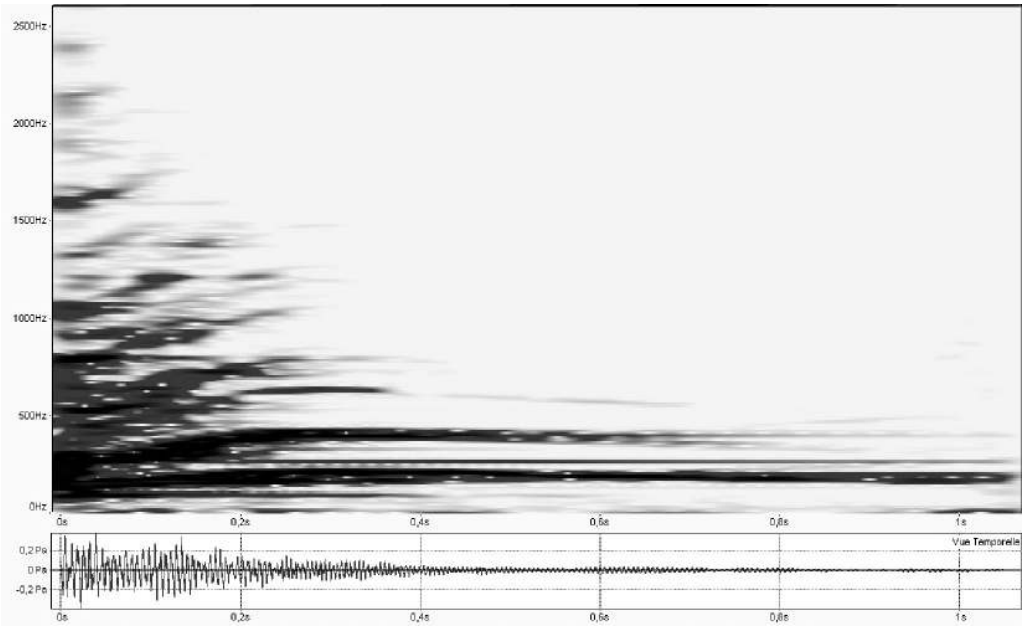

Figure 7. Spectrogramme d'un son entretenu à hauteur variée (Y selon la classification de Schaeffer)

Conclusion

Dans cette étude, des mots ont été associés à des objets sonores et présentés à des sujets dans le but d'aborder la notion de sémiotique des objets sonores du point de vue cognitif et acoustique.

Nous avons ainsi montré que l'association mots/sons est souple car plusieurs mots ont été associés à chaque son lors de la classification libre dans la première partie de l'expérience. Malgré cette constatation, notre proposition de couples mots/sons associés a dans l'ensemble semblé convenir aux participants de l'expérience, puisque la variation inter-individuelle est faible (taux d'erreur de 16%). Lorsque les couples mots/sons non associés sont présentés aux sujets, le taux d'erreur est encore plus faible (de l'ordre de 5%). Les mesures électrophysiologiques ont permis de montrer que le traitement d'un son est influencé par le mot qui le précède et que ce traitement semble similaire au traitement de la sémantique dans le langage. Ceci pourrait indiquer que le processus d'attribution du sens pourrait être commun au langage et aux objets sonores.

Du point de vue acoustique, les premières analyses indiquent un lien possible entre paramètres acoustiques et description sémantique des sons. Ainsi, les variations de fréquence seraient liées à une impression de mouvement. Les sons à hauteur variable sont souvent caractérisés par des mots tels que rebond, monté, roulement, rotation, etc. La densité spectrale est un autre paramètre acoustique qui semble avoir une forte influence sur l'appréciation des sons. Ainsi, un son caractérisé par une densité spectrale large bande et un comportement stationnaire est généralement associé à des situations désagréables, évoquant des mots tels que frisson, froid, douleur, etc. Ces analyses devront être approfondies de façon à identifier les éventuels descripteurs de timbre (barycentre spectral, temps d'attaque, flux spectral, etc.) qui sont spécifiquement liés aux descriptions verbales.

Remerciements

Ce projet a en partie été financé par l'Agence Nationale de la Recherche dans le cadre du projet "senSons – vers le sens des sons" (http ://www.sensons.cnrs-mrs.fr)

Bibliographie

ARAMAKI, M. & KRONLAND-MARTINET, R. (2006), « Analysis-Synthesis of Impact Sounds by Real-Time Dynamic Filtering », IEEE *Transactions on Audio, Speech, and Language Processing*, Vol. 14, No. 2, 695-705.

BERGER, H. (1929), « Über das Electrenkephalogramm das Menchen », *Archiv für Psychiatrie*, 87, 527-570.

BENSA, J., BILBAO, S., KRONLAND-MARTINET, R. & SMITH, J.O. (2003), « The Simulation of Piano String Vibration : from Physical Model to Finite Difference Schemes and Digital Waveguides », *Journal of the Acoustical Society of America*, Vol. 114, 1095-1107.

BENSA, J., DUBOIS, D., KRONLAND-MARTINET, R. & YSTAD S. (2005), « Perceptive and Cognitive Evaluation of a Piano Synthesis Model », *Lecture Notes in Computer Sciences (LNCS)*, Springer Verlag, 232-245.

BENTIN, S., MCCARTHY G. & WOOD, C. (1985), « Event Related Potentials, Lexical Decision, and Semantic Priming », *Electroencephalography and Clinical Neuropsychology*, 60, 353-355.

BUXTON, R.B. (2002), *Introduction to Functional Magnetic Resonance Imaging.* Cambrige, UK : Cambridge University Press,

GAVER, W. (1993), « What in the World Do We Hear ? An Ecological Approach to Auditory Source Perception », *Ecological Psychology*, 5(1), 1-29.

KOELSCH, S, , KASPER, E., SAMMLER, D., SCHULZE, K., GUNTER, T. & FRIEDERICI, A.D. (2004), « Music, Language and Meaning : Brain Signatures of Semantic Processing », *Nature neuroscience*, vol. 7, n° 3.

KUTAS, M. & HILLYARD, S.A. (1980), « Reading Senseless Sentences : Brain Potentials reflect Semantic Incongruity », *Science*, vol. 207, 203-204.

OVERY, K. (2003), « Dyslexia and Music : From Timing Deficits to Musical Intervention », *Ann. N.Y. Acad. Sci.*, 999, 497-505.

PATEL, A., GIBSON, E., RATNER, J., BESSON, M. & HOLCOMB, P. (1998), « Processing Syntactic Relations in Language and Music : An Event-Related Potential Study », *Journal of Cognitive Neuroscience*, 10, 717-733.

RUGG et al. (1995) ; RUGG, M. & COLES, M.G.H. (1995), « Electrophysiology of Mind : Event Related Brain Potentials and Cognition », *Oxford Psychology Series*, 25.

SCHAEFFER, P. (1966), *Traité des Objets Musicaux*, Paris : Éditions du Seuil.

SCHÖN, D., MAGNE, C., & BESSON, M. (2004), « The Music of Speech : Electrophysiological Study of Pitch Perception in Language and Music », *Psychophysiology*, 41, 341-349.

Annexe

Liste des mots (associés et non associés) utilisés dans l'expérience :

N12	angoisse	léger	N9	magie	souffrance
Y''9	passage	massif	Y'2	massif	frottement
Y13	appel	bulle	X4	mystique	électrique
Y'3	bulle	plainte	X''3	message	mystique
X'5	coupe	secret	Y6	montée	suspense
X''9	crissement	glisse	X5	peur	dérapage
Y8	dérapage	délice	Y17	plainte	univers
Y''5	descente	géant	X''1	plongée	peur
Y14	douleur	frappe	Y4	rayure	lointain
X''8	électrique	rebond	Y15	rebond	énervé
Y22	énervé	coupe	X7	rotation	tension
X9	foule	squelette	Y1	roulement	froid
N'5	frappe	crissement	X8	secret	angoisse
Y19	frisson	roulement	X'7	souffle	magie
X15	froid	joie	N4	souffrance	vitesse
Y23	frottement	descente	Y5	squelette	frisson
Y7	géant	montée	Y25	délice	douleur
Y12	glisse	foule	N6	suspense	rotation
X''10	grenier	souffle	X2	ténèbres	appel
Y3	guerre	grenier	N2	tension	passage
Y9	joie	guerre	Y18	univers	ténèbres
Y21	léger	plongée	Y'4	vitesse	message
X13	lointain	rayure			