



HAL
open science

Cognition spatiale auditive

Alain Cheveigné (de)

► **To cite this version:**

| Alain Cheveigné (de). Cognition spatiale auditive. 2005. hal-00003507

HAL Id: hal-00003507

<https://hal.science/hal-00003507>

Preprint submitted on 20 Jan 2005

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.

Cognition spatiale auditive

Responsable scientifique : Alain de CHEVEIGNÉ

Alain de CHEVEIGNÉ
IRCAM
1 place Igor Stravinsky
75004
01 44 78 48 46
cheveign@ircam.fr

Sous-thèmes dont relève ce projet :
Espace
Audition
Sons

Équipes partenaires :

- Équipe Acoustique des Salles (Ircam, 1 place Igor Stravinsky, 75004 – Tél : 01 44 78 48 85)

Résumé signalétique

La *cognition spatiale auditive* couvre l'apport des indices sonores à l'organisation cognitive de l'espace, et l'apport des indices spatiaux pour l'organisation perceptive des scènes sonores complexes. La question a été abordée du point de vue de la théorie et de la modélisation, de l'expérimentation psychoacoustique, et des applications. Un modèle nouveau («Réseau de Corrélation») a été développé qui permet d'implémenter de nombreuses fonctions de traitement binaural (localisation, ségrégation) et monaural (hauteur, timbre, ségrégation) de façon unifiée et proche de la physiologie. Il constitue aussi une base utile pour le développement de techniques de traitement du signal. Une série d'expériences a permis d'explorer certains indices spatiaux qui déterminent l'intelligibilité de la parole, et l'influence sur ces indices des techniques de restitution sonore. Des applications ont été développées dans le cadre des activités de normalisation MPEG-7, et plusieurs brevets ont été déposés sur la base de ces travaux. Ils concernent notamment des méthodes performantes d'estimation de fréquence fondamentale (simple ou multiple) directement inspirées du modèle de réseau de corrélation, ainsi que des techniques de description multimédia.

Mots-clés : Espace • perception spatiale • cognition spatiale • ségrégation • analyse de scène auditive • modélisation • traitement du signal • estimation spatiale • estimation spectrale • estimation de hauteur • perception de la parole • indexation multimédia.

Participants :

Chercheurs : 3

Thésards : 3

Ingénieur : 1

Nombre de hommes-mois: 18

Rappel des enjeux et objectifs fixés à l'origine

Le projet s'intéresse à la formation de représentations spatiales à partir d'informations acoustiques, et aux interactions entre indices acoustiques, visuels et cognitifs. Sont plus particulièrement visés les mécanismes qui permettent la coexistence perceptive de sources sonores multiples dans une scène auditive complexe. Les facteurs sont de plusieurs ordres: (1) caractéristiques d'une source cible qui favorisent sa saillance, (2) caractéristiques des sources concurrentes qui favorisent leur «transparence», (3) propriétés de transmission (acoustique des salles, etc.) qui modifient ces caractéristiques, (4) corrélations avec d'autres modalités (vision), (5) apports de modèles spatiaux de niveau supérieur, attentes, etc., (6) interactions entre action et perception.

Les objectifs scientifiques sont:

1. Une meilleure compréhension des mécanismes d'organisation spatiale auditive. Sont visés les mécanismes neuronaux, les rapports entre localisation et ségrégation, et ceux entre indices binauraux et d'harmonicité.
2. Une meilleure connaissance des divers facteurs multimodaux et cognitifs qui interagissent avec les facteurs acous-

tiques dans l'élaboration des représentations spatiales. Ces facteurs peuvent inclure aussi les stratégies actives, celles de l'individu qui perçoit comme celles de celui (locuteur ou musicien) qui donne à percevoir.

3. Le développement d'une «mesure de transparence» qui caractérise ces facteurs de manière quantitative, et qui permette de prédire les effets de la juxtaposition de plusieurs sources sur leur saillance perceptive. Cette mesure prédirait la limite de la complexité perceptible d'un ensemble de sons en fonction de ses éléments, de facteurs négatifs tels que l'acoustique de la salle, et de facteurs positifs tels que corrélats visuels, connaissance de la partition, etc..

Des applications sont possibles dans les domaines de l'aide à la conception acoustique de systèmes ou bâtiments, composition musicale, conception de sons (lutherie ou «design sonore»), signalétique sonore et systèmes multimodaux de présentation de données complexes, appareillages pour mal-entendants, systèmes de télécommunications (téléphones statifs, téléconférence), systèmes de communication homme-machine, etc.

Résumé des résultats effectivement atteints

Parmi les *objectifs scientifiques* de départ, l'objectif (1) a été atteint. Le modèle Réseau de Corrélation (RC) constitue un modèle utile pour la compréhension des mécanismes fondamentaux d'organisation spatiale auditive. Il explique comment certaines fonctions peuvent être assurées, et facilite la mise en correspondance des modèles fonctionnels avec les mécanismes et structures neuronaux. Il unifie dans un cadre commun des modèles de localisation, hauteur, et ségrégation sur la base d'indices binauraux et d'harmonicité.

L'objectif (3) a été abordé sous l'angle de la valorisation dans le cadre du processus d'élaboration de la norme MPEG-7. Cependant les contraintes et logique propres au processus MPEG-7 nous ont fait dévier de l'objectif d'origine, qui était l'inscription dans la norme d'une «mesure de transparence», vers d'autres objectifs qui ont été atteints mais dont les liens

avec le projet sont plus ténus. L'objectif (3) a aussi été abordé dans les expériences qui ont mesuré les effets des indices acoustiques spatiaux (produits par simulation) sur l'intelligibilité.

L'objectif (2) a été abordé dans la phase d'analyse bibliographique uniquement.

Le projet de départ se fixait des *objectifs de valorisation*. Ceux-ci ont été atteints de façon tangente mais efficace. Quatre brevets ont été déposés, dont deux issus des activités MPEG-7 et deux du modèle RC (un cinquième brevet issu de ce modèle est en cours d'élaboration). En particulier, une solution efficace a été apportée au problème de l'estimation de F0, important dans les applications musicales et de traitement de la parole.

Publications issues du projet

Brevets

- de Cheveigné, A. (2001), «Method for the scaling of the indexing data of a multimedia document,» Brevet US déposé 26 Sept 2001.
- de Cheveigné, A. (2001), «Méthode d'indexation dans une structure de description multimédia,» Brevet France déposé 30 Apr 2001, no 01 05893.
- de Cheveigné, A. (2001), «Estimation de périodes fondamentales de sources concurrentes multiples notamment de son,» Brevet France déposé 17 Mai 2001, no 01 06683.
- de Cheveigné, A. (2001), «Procédé d'extraction de la fréquence fondamentale d'un signal sonore au moyen d'un dispositif mettant en oeuvre un algorithme d'autocorrélation,» Brevet France déposé 01/06/2001, no 01 07284.

Articles parus ou à paraître

- de Cheveigné, A., and Kawahara, H. (2002). «YIN, a fundamental frequency estimator for speech and music,» J. Acoust. Soc. Am. à paraître.
- de Cheveigné, A., and Kawahara, H. (2001). «Comparative evaluation of F0 estimation algorithms,» Proc. Eurospeech, 2451-2454.
- [10] de Cheveigné, A., and Pressnitzer, D. (2001). «Modèle Réseau de Corrélation de traitement auditif,» La cognition: de l'artificiel au naturel, Paris, Collège de France, p13.

Documents techniques

- de Cheveigné, A., and Smith, B. (1999), «A 'Sound Transparency' descriptor,» ISO/IEC JTC1/SC29/WG11, MPEG99/m5198.

- de Cheveigné, A., and Peeters, G. (1999), «Scale tree,» ISO/IEC JTC1/SC29/WG11, MPEG99/m5076.
- de Cheveigné, A. (1999), «Scale tree update,» ISO/IEC JTC1/SC29/WG11, MPEG99/m5443.
- de Cheveigné, A., and Peeters, G. (1999), «A set of scalable audio descriptors,» ISO/IEC JTC1/SC29/WG11, MPEG99/m5444.
- de Cheveigné, A. (2000), «Core experiment on the scale tree,» ISO/IEC JTC1/SC29/WG11, MPEG00/m5910.
- de Cheveigné, A., and Peeters, G. (2000), «Core set of audio signal descriptors,» ISO/IEC JTC1/SC29/WG11, MPEG00/m5885.
- de Cheveigné, A., Pfeiffer, S., and Casey, M. (2000), «Framework for audio descriptors,» ISO/IEC JTC1/SC29/WG11 N3242.

Logiciels

logiciel YIN d'estimation de F0, écrit en Matlab, mis à disposition des chercheurs à fins de recherche non-lucratives exclusivement, <http://www.ircam.fr/pcm/cheveign/sw/yin.zip>, <http://www.ircam.fr/pcm/cheveign/sw/sf.zip>.

Mémoire de fin de stage

Gretzki, R. (2002) «Spatial segregation and intelligibility of concurrent speech», Studienarbeit report, Univ. Bochum.

