

Fiche descriptive de la thèse

Campagne 2013

Encadrant de la thèse Orange Labs : Rozenn NICOL

Pôle : OLPS

Direction : COMSERV

Département : SVQ

Projet : Technologies & Devices

Terrain de conquête : TC CED

Site : Lannion

Sujet de la thèse :

Application des processus perceptifs d'Analyse de Scène Auditive à la représentation et la compression d'une scène sonore complexe

Contexte global de l'étude et état de l'art

En perception sonore, le domaine de l'Analyse de Scène Auditive (Auditory Scene Analysis [1]) décrit l'ensemble des processus perceptifs et cognitifs impliqués dans la perception d'une scène sonore et par lesquels l'auditeur établit une hiérarchisation de l'information sonore. Ainsi l'auditeur ne perçoit pas (du moins consciemment) tous les détails de la scène. On observe par exemple des phénomènes de masquage dans lesquels des informations (voire des sources sonores) disparaissent au profit d'autres. La précision de localisation sonore peut aussi être affectée. Enfin, on observe une réorganisation de la scène sonore selon des flux auditifs, dans lesquels des sources sonores sont fusionnées en différents groupes en fonction de leurs propriétés temporelles, fréquentielles et spatiales. Une partie des processus est connue, mais il reste encore de nombreux aspects à clarifier.

L'objectif de la thèse est de développer un modèle basé sur ces processus perceptifs d'analyse de scène auditive et permettant de prédire à partir d'un flux audio multicanal quelconque (quel que soit son format de représentation audio 3D), quelle est l'information utile, i.e. contribuant effectivement à la perception de l'auditeur. Ce modèle pourra être alors mis en œuvre pour optimiser la compression des futurs formats audio 3D (par ex : 22.2, Higher Order Ambisonics), pour lesquels le nombre de canaux (plusieurs dizaines) requièrent des débits prohibitifs pour une diffusion sur les réseaux grand public. Plus généralement le modèle pourrait être appliqué à tout format de représentation audio 3D. Le modèle permettrait aussi d'enrichir des algorithmes de séparation de sources. Enfin, il pourrait aussi enrichir les méthodologies d'évaluation objective et subjective du son spatialisé.

Le premier verrou scientifique à lever est la compréhension et la modélisation de la perception d'une scène sonore complexe. Cette perception implique des processus bas et haut niveau dont beaucoup restent à identifier.

Le second verrou est de tirer parti de ces processus, en exploitant les limitations de la perception, pour optimiser la représentation et la compression des contenus audio 3D.

Concernant la compression, on constate aujourd'hui 2 lacunes :

- Pour l'instant, les codecs développés ne s'appliquent qu'aux contenus stéréo ou 5.1. Il manque des schémas de compression applicables aux futurs formats audio multicanaux composés de plusieurs dizaines de canaux (22.2, Higher Order Ambisonics).
- Mais surtout il y a peu ou pas de prise en compte de la perception audio spatialisée (bas et surtout haut niveau) dans les schémas de compression proposés actuellement.

[1] Bregman, A. S. (1990) "Auditory scene analysis", MIT Press: Cambridge, MA.

[2] Gallo, E. (2006) « Restitution sonore hiérarchique et perceptive d'environnements virtuels multimodaux », PhD Thesis, Univ. de Nice-Sophia Antipolis.

[3] Tsingos, N. (2007) "Perceptually-based auralization", 19th ICA, Madrid.

[4] Bidet-Caulet, A. (2007) « Mécanismes neurophysiologiques de la perception des flux sonores chez l'Homme : effets des contextes acoustiques et attentionnels », PhD Thesis, Univ. Claude Bernard Lyon 1.

Objectifs de la thèse / Résultats attendus / Défis scientifiques et techniques à relever

Le principal objectif de la thèse est de construire un modèle reproduisant les processus perceptifs d'Analyse de Scène Auditive et permettant à partir d'un flux audio multicanal quelconque de prédire quelle est l'information perceptivement utile.

La première étape passera par des expérimentations pour préciser les processus perceptifs et cognitifs mis en œuvre dans la perception d'une scène sonore complexe, afin de compléter les connaissances antérieures.

Une question fondamentale est de clarifier le poids relatif des informations fréquentielles et spatiales dans le jugement de perception. Des études suggèrent en effet que, dans certains cas, l'information fréquentielle prend le pas sur l'information spatiale, ce qui affecte la position spatiale perçue des sources, voire conduit à des regroupements de plusieurs sources. Ce type de phénomènes ouvre des perspectives prometteuses en termes de compression.

Une autre question fondamentale traitée par la thèse est celle du format de représentation audio 3D. Le modèle devra accepter en entrée n'importe quel format audio 3D : il sera donc nécessaire d'une part d'identifier un format audio 3D générique, et d'autre part de développer des outils de conversion pour passer d'un format à l'autre. Enfin, l'objectif ultime de la thèse est de proposer un nouveau format audio 3D basé sur la perception spatiale auditive.

Approche méthodologique proposée par le responsable technique

Le travail de thèse comprendra 2 volets :

- Une première phase d'expérimentations psycho-acoustiques afin de compléter les connaissances de l'état de l'art, notamment sur la question des interactions entre les propriétés fréquentielles et spatiales des sources d'une scène sonore complexe,
- Une seconde phase de modélisation des processus d'Analyse de Scène Auditive, pour produire un algorithme capable de prédire la scène sonore perçue à partir d'un flux audio 3D.

Les compétences nécessaires à ces travaux relèvent essentiellement de la psycho-acoustique, de la perception spatiale auditive, de la méthode expérimentale, et plus spécifiquement de l'Analyse de Scène Auditive.

Contact

Rozenn Nicol
rozenn.nicol@orange.com
02 96 05 16 99