

Proposition d'un sujet de post-doc:

Formats audio multicanal, conversion de formats et compression

Contexte :

Actuellement, la diffusion et le transport des contenus audio multi-canal (donc immersif) reposent essentiellement sur le format standard 5.1 dédié à une restitution pour un dispositif de 5 haut-parleurs (disposition normalisée par l'ITU). Il existe pour les contenus de ce type une palette de codecs adressant différentes gammes de qualité et débit (du AC3, et HE-AAC multicanal, au MPEG Surround basé sur un principe de codage spatial paramétrique). Les contenus concernés sont à l'origine associés à des modes de consommation et de production très particuliers : consommation "passive" par le grand public de contenu transmis de façon monolithique et généré par des studios professionnels (cinéma, télévision....).

Aujourd'hui de nouveaux formats audio multicanal sont proposés avec une certaine surenchère du nombre de canaux (déclinaisons 6.1, 7.1, 10.2, 22.2 du format 5.1). Une autre innovation mise en avant est un enrichissement des modèles d'analyse et de représentation de la scène sonore (formats WFS (Wave Field Synthesis), HOA (Higher Order Ambisonics), ou binaural).

Problème :

Dans ce contexte, les enjeux et questions qui émergent sont :

- Quel est l'apport de ces nouveaux formats? Quels sont leurs atouts pour prendre leur place dans le paysage audio multicanal, notamment dans un contexte largement dominé par le format 5.1? En ont-ils la réelle capacité? Quelles sont leurs faiblesses?
- La percée des formats avec un nombre accru de canaux s'accompagne nécessairement du développement d'outils de conversion de formats, permettant notamment de réduire (techniques de downmix) ou d'augmenter (techniques de upmix) le nombre de canaux en préservant la qualité spatiale, voire en l'améliorant dans le cas du upmix, ceci d'une part pour assurer une compatibilité des nouveaux formats avec les équipements existants et d'autre part pour réaliser la promotion des nouveaux formats en créant des contenus enrichis à partir des contenus existants.
- Pour les mêmes raisons, des outils de transcodage permettant de passer d'un modèle de représentation à un autre (par exemple conversion d'un flux 5.1 en flux WFS) sont aussi à prévoir.
- Dans la perspective de transporter et stocker les contenus multicanaux, des outils de compression sont nécessaires. Les schémas existants, développés pour les flux stéréo et 5.1 (compression audio 3D de type paramétrique: BCC, MPEG Surround...), présentent des limites, aujourd'hui plus ou moins identifiées, d'une part face à l'inflation du nombre de canaux, et d'autre part parce qu'ils sont inadaptés aux nouveaux modes de représentation audio 3D.
- Les techniques d'analyse de scène sonore sont des outils prometteurs pour développer de nouveaux outils de conversion et de compression, dans la mesure où elles permettent d'aborder le problème de façon globale et générique, indépendamment d'un format ou d'un modèle de représentation.
- Il faut aussi prendre en compte les connaissances sur la perception audio spatialisée, ainsi que les mécanismes d'analyse de scène auditive.
- Il faut également avoir conscience de l'évolution des contextes d'utilisation de l'audio multicanal qui aujourd'hui n'est plus limité à la production musicale ou cinématographique, mais commence à conquérir de nouveaux horizons tels que la communication (conférence audio spatialisée, partage d'ambiance), les jeux ou les contenus autoproduits.
- A cette évolution des contextes s'ajoute une évolution des usages d'une consommation passive des contenus vers une consommation de plus en plus interactive (exemples de scénarii: modification de contenu par l'utilisateur pour les conférences audio immersives, ajustement séparé commentaires/ ambiance pour la diffusion de matchs de foot, karaoké) que doivent permettre les nouveaux formats, comme l'atteste par exemple l'activité de normalisation autour du SAOC (Spatial Audio Object Coding) qui aboutit courant 2009.

Objectifs du postdoc :

Le travail attendu consistera en premier lieu à établir un panorama de l'audio multicanal autour des points suivants :

- état des lieux des formats audio multicanaux existants, reconnus ou pressentis,
- positionnement des nouveaux formats par rapport aux formats existants (évaluer leur apport, leurs forces et leurs faiblesses)

- identification des limites des schémas de compression standards pour les nouveaux formats

Dans une seconde étape, sur la base de ce panorama, il s'agira de:

- définir des scénarii privilégiés d'exploitation des nouveaux formats,
- pour ces configurations développer des outils de conversion et de compression, notamment sur la base des techniques d'analyse de scène sonore (techniques d'identification/ localisation/ séparation de sources).

Une piste prometteuse qui sera étudiée en particulier sera l'utilisation des techniques de upmix / downmix dans la chaîne audio dans un but de compression. Par exemple, le développement d'algorithmes d'upmix pour le format HOA permet d'utiliser à la prise de son un dispositif ne comportant que 3 ou 4 capsules et par suite un aussi faible nombre de canaux à transmettre, tout en offrant la possibilité d'une restitution avec une qualité de spatialisation étendue aux ordres supérieurs.

Profil attendu / compétences recherchées ou appréciées :

- compétences générales: traitement du signal; acoustique
- domaine audio 3D; de préférence expérience / connaissance des formats audio 3D
- expérience souhaitée dans au moins l'un des domaines suivants:
 - o codage audio spatial paramétrique
 - o upmix
 - o séparation de source
- maîtrise de la programmation C / C++ et matlab

Durée et localisation :

- 12 mois renouvelables 6 mois
- Orange Labs, site de Lannion (22)

Contact :

Jérôme Daniel
Orange Labs / TECH / SSTP
2, avenue Pierre Marzin, 22307 Lannion Cedex
e-mail: jerome.daniel@orange-ftgroup.com
tél: +33 2 96 05 27 96

Bibliographie indicative :

- AES 19th International Conference: **Surround Sound - Techniques, Technology, and Perception, Schloss Elmau, Germany, 2001 June 21-24**
- AES anthology: **SPATIAL SOUND TECHNIQUES PART 2 - Multichannel Audio Technologies** edited by Francis Rumsey
- **Spatial Audio Object Coding (SAOC) – The Upcoming MPEG Standard on Parametric Object Based Audio Coding**. Presented at the 124th AES Convention 2008 May 17–20 Amsterdam, The Netherlands
- V. Pulkki et al (DirAC)
 - o V. Pulkki "Directional audio coding in spatial sound reproduction and stereo upmixing". To appear in AES 28th Int. Conf. Pitea, Sweden, June 2006
 - o V. Pulkki and C. Faller "Directional audio coding: Filterbank and STFT-based design. In 120th AES Convention, Paris, France, May 20-23, 2006. Audio Engineering Society. Paper # 6658
- Jot, Goodwin, Avendano:
 - o J.-M. Jot, J. Merimaa, M. Goodwin, A. Krishnaswamy, and J. Laroche. **Spatial audio scene coding in a universal two-channel 3-D stereo format**. Convention of the Audio Engineering Society, October 2007.
 - o M. Goodwin. **Enhanced microphone-array beamforming based on frequency-domain spatial analysis-synthesis**. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, October 2007.
 - o M. Goodwin and J.-M. Jot. **Multichannel surround format conversion and generalized upmix**. AES 30th International Conference. March 2007. Saariselkä, Finland
- Higher Order Ambisonics:
 - o Daniel, J., R. Nicol, and S. Moreau. **Further Investigations of High Order Ambisonics and Wavefield Synthesis for Holophonic Sound Imaging**. in *114th AES Conv.* 2003. Amsterdam, The Netherlands.
 - o Moreau, S., J. Daniel, and S. Bertet. **3D Sound Field Recording with Higher Order Ambisonics - Objective Measurements and Validation of a 4th Order Spherical Microphone**. in *AES 120th Convention*. 2006. Paris, France.