Improved audio matting and re-rendering from field recordings.

Supervisor: N. Tsingos
REVES – INRIA Sophia Antipolis
http://www-sop.inria.fr/reves/
Nicolas.Tsingos@sophia.inria.fr

We recently proposed a novel approach to real-time spatial rendering of realistic auditory environments and sound sources recorded live, in the field [1,2]. Using a set of standard microphones distributed throughout a real-world environment we record the sound-field simultaneously from several locations. After spatial calibration, we segment from this set of recordings a number of auditory components, together with their location. We compared existing time-delay of arrival estimations techniques between pairs of widely-spaced microphones and introduced a novel efficient hierarchical localization algorithm. Using the high-level representation thus obtained, we can edit and re-render the acquired auditory scene over a variety of listening setups. In particular, we can move or alter the different sound sources and arbitrarily choose the listening position. We can also composite elements of different scenes together in a spatially consistent way. Our approach provides efficient rendering of complex soundscapes which would be challenging to model using discrete point sources and traditional virtual acoustics techniques.
To deal with scenes containing a large amount of stationary background noise, we propose to segment out a background and foreground component which can be processed and re-rendered separately.
Example results are available at http://www-sop.inria.fr/reves/projects/audioMatting and http://www-sop.inria.fr/reves/projects/aes30.

In this project, we propose to explore several ways of improving this matting and re-rendering technique. We propose to improve on subband localization by exploring techniques used in computer vision such as RANSAC [3]. Instead of computing a 3D position for each subband at regular time-intervals, we propose to define and localize a large set of “good” features for which we are guaranteed to obtain a good correspondence between recordings, similar to SIFT features in computer vision [4]. For instance, we could try exploiting such features on an equivalent time-frequency domain image of the sound or define a novel set of features better designed for the task of maximizing the quality of matching between signals. This could be used to solve both for recording setup configuration and position of the subbands simultaneously.
Another direction is to explore how sparser representations obtained by decomposing the signals on a overcomplete set of basis functions [5] could be used to improve both the representation, encoding and matching between the recordings.
Real-time implementation of a calibration and matching mechanism, for instance by exploiting processing power of graphics processors (GPUs), would also be mandatory to address live/stage applications.
Our approach is currently being evaluated by French car manufacturer Renault for sound design purposes. The project would also involve testing using data sets provided by Renault.

References

[1] Emmanuel Gallo, Nicolas Tsingos and Guillaume Lemaitre. 3D audio matting, editing and re-rendering from field recording. EURASIP JASP, special issue on spatial sound and virtual acoustics, 2007.
[2] Emmanuel Gallo and Nicolas Tsingos. Extracting and Re-rendering Structured Auditory Scenes from Field Recordings. AES 30th Intl. Conference on Intelligent Audio Environments, 2007.

[3] M.A. Fischler and R.C. Bolles. Random Sample Consensus: A Paradigm for model fitting with applications to image analysis and automayed cartography.

[4] D.G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints.