Spatial perception of virtual 3D sound scene: how to assess the quality of 3D audio rendering by WFS?

Rozenn Nicol[1], Laetitia Gros[1]

[1] *France Telecom R&D, Lannion, France, Email: rozenn.nicol@rd.francetelecom.com, laetitia.gros@rd.francetelecom.com*

Introduction

In the context of the IST Project CARROUSO [5], listening tests are carried on by France Telecom R&D, in order to assess the quality of 3D-audio rendering provided by WFS. The experiment is focussed on WFS rendering, which will be compared with other sound spatialization technologies, such as Ambisonics, intensity panning and 5.1. All these systems are based on 3D-audio rendering provided by loudspeaker array of more or less complexity and the various experiments use a common 3D-audio demonstrator. In order to assess the quality of the 3D-audio rendering, two different, but complementary, aspects will be analyzed: first, the localization accuracy provided by the system, second, the overall quality, through different criteria, such as intelligibility, realism or immersion. This paper deals with a first experiment, which consists in a localization test comparing WFS and Ambisonics.

3D audio rendering by WFS and Ambisonics

Recent work at France Telecom R&D has shown how WFS and Ambisonics are close technologies of sound spatialization [1]. However, it should be noticed that Ambisonics, here, does not refer to *standard Ambisonics*, i.e. first order Ambisonics, but to Ambisonics generalized to higher orders, i.e. *High Order Ambisonics* (HOA). Then, WFS [2] and HOA [3] provide equivalent solutions of 3D audio rendering [1]. First, they are both based on loudspeaker array. Second, they are both able to synthesize *spherical* waves (i.e. close sources), as well as *plane* waves (i.e. far sources), which means the ability to render the depth of the sound scene with varying finite distance sources. Thirdly, they are both able to synthesize *outside* sources (i.e. sources located beyond the loudspeakers from the listener point of view), as well as *inside* sources (i.e. sources located between the listener and the loudspeakers). For WFS, synthesizing inside sources is based on time reversal, whereas for HOA, Nearfield Compensation Filters [1] are used.

Therefore, our first objective is to perceptually evaluate and compare 3D audio rendering by WFS and HOA. It should be remarked that, to our knowledge, this experiment is the first subjective evaluation of HOA. Since WFS and HOA both aim at physical reconstruction of the acoustic waves, a *localization test*, which is a usual methodology used for the evaluation of binaural synthesis for instance, has been judged as particularly relevant for this first subjective assessment experiment focused on WFS and HOA. Indeed, contrary to other sound spatialization technologies, such as stereophony or 5.1, which aim at rather conveying an overall sound scene to the listener, than giving him an identical copy of each details of the sound scene, WFS and HOA are potentially able to reproduce the full details of each sound source, as for its temporal and spatial properties. A localization test thus intends to verify this ability. To some extent, WFS and HOA may be considered, like binaural synthesis, as "research laboratory" technologies, which are more dedicated to virtual reality context than to the recording and reproduction of sound event, and thus less recognized by the sound engineer community. Nevertheless, listening tests, which are a more usual way of perceptually assess sound spatialization, will be performed in a second step. For this second experiment, WFS and HOA will be evaluated in comparison with "more standard" sound spatialization technologies, such as 5.1 and intensity panning. For the localization experiment, however, it has no sense comparing WFS and HOA with other sound spatialization technologies, since they are currently the two only methods being able to render finite distance and enclosed sources.

Localization test

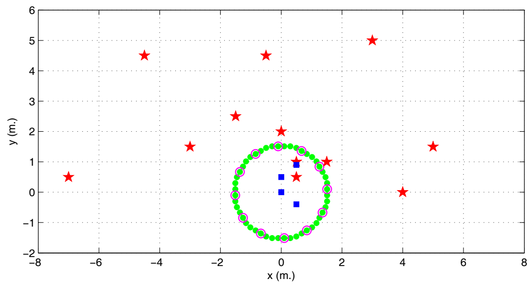

Figure 1: Geometrical layout: Description of the loudspeaker array (green circles: WFS, magenta circles: HOA), the virtual sound source positions (red stars) and the listening positions (blue squares).

2D-audio demonstrator

For this test, only the horizontal plane is considered. The experiment is performed with a 2D-audio demonstrator, which consists of a 48-loudspeaker dodecagonal horizontal array (Figure 1). An acoustically transparent curtain hides the loudspeakers, in order to remove the influence of the visual modality on the auditory one. The loudspeakers are fed by one PC unit, which is equipped by two 24-channel sound cards. All the spatialization effects are created by the FTR&D software library, which implements both WFS and HOA rendering. WFS rendering uses all the 48 loudspeakers, whereas HOA rendering is based on fifth order Ambisonics, which is decoded for a 12-loudspeaker regular array (Figure 1).

Methodology

The localization of 12 virtual sound sources (Figure 1) is tested, with various azimuth and distance (randomly distributed between 0 and 180° for the azimuth between 0.5 and 7 m. for the distance). The localization is limited to the frontal plane, considering the symmetry of rendering of the WFS and HOA systems. Besides, 4 listening positions (Figure 1) are considered, in order to evaluate the audio rendering at both centered and non centered locations. These latter may be in the front or in the rear [1]. The influence of the neighborhood of the loudspeakers is also studied.

Two kinds of stimuli are considered: a high-pass filtered white noise (500 ms duration noise burst, played 3 times) and speech samples. The first one is used to specifically study the aliasing effect on the localization accuracy and to compare it for the two methods. Since the loudspeaker spacing is 20 cm, it is expected that spatial aliasing occurs for frequencies greater than 850 Hz, but only for WFS rendering. Contrary to WFS, HOA rendering is free from spatial aliasing [1]. Speech samples that present a wideband spectrum shall not put in evidence any aliasing effect and is used to compare the two systems for natural sounds. Two male and two female speakers are considered, with twelve samples per speaker (3 second duration).

Localization Subject Reporting

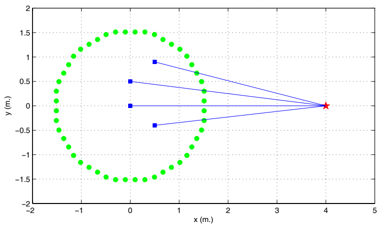

Figure 2: Perceived distance valuation by crossing the subject pointings reported for the different listening positions.

A head-tracker is used for the localization judgement reporting. The subject, equipped with the head-tracker on his/her head, is located at one of the four listening positions marked on the ground. For each stimulus, the subject is asked to point his/her head into the direction from which he/she perceives the sound. Once the stimulus is localized, the subject has to keep the angle during three seconds, validating in this way the perceived direction. This method, intuitive for the subjects, also enables to measure the perceived distance [4]. Indeed, since each virtual sound source is localized from 4 listening positions, it is possible to draw a line from each listening point according to the perceived direction. Thus, the crossing "point" (in practice rather an area than a point) of the 4 lines gives an approximation of the perceived location of the virtual sound source, not only in azimuth but also in distance, as depicted by Figure 2.

Test Protocol

The test is composed of two 1-hour sessions, with one session per stimulus. For each session, and for each listening point, each virtual sound source is presented eight times in all, four times rendered by WFS and four times by HOA. Therefore, all in all, for each listening point, 96 stimuli (12 sound source position x 2 methods x 4 presentations) are presented in a random order, which differ for each subject. Then this sequence is repeated for the three other listening positions. The listening position order is random from one subject to another, in order to compensate the order effects. In addition, half the subjects begin the test with the high-filtered noise session, whereas the other half begin with the speech session. Sixteen adult subjects, eight experts and eight naïves, perform the test.

Conclusion and future work

The methodology of a localization test comparing the WFS and HOA rendering has been described. Some preliminary results of the experiment will be shown at the oral presentation. Future work will focus on quality assessment of 3D-audio rendering, through three main experiments: first, a speech intelligibility test, in order to evaluate the improvement provided by WFS rendering in comparison with other sound spatialization technologies, second, a test based on the collection of multi criteria judgment, in terms of naturalness, space impression, envelopment, accuracy and agreement. The third experiment deals with the 3D-audio immersion and consists in evaluating how a subject feels involved into a virtual complex sound scene. By this way, it is also intended to derive innovative and comprehensive test methodologies in order to assess the perceptual aspects of 3D audio system rendering. For these experiments, WFS is compared to various sound spatialization technologies: not only to HOA, but also 5.1 and intensity panning.

References

[1] **J. Daniel, R. Nicol, and S. Moreau.** *Further Investigations of High Order Ambisonics and Wavefield Synthesis for Holophonic Sound Imaging.* In AES 114th Convention. Amsterdam (2003)

[2] **R. Nicol.** *Restitution Sonore Spatialisée sur une Zone Etendue: Application à la Téléprésence.* Université du Maine: Le Mans, France (1999), http://gyronymo.free.fr/audio3D/Guests/RozennNicol_PhD.html

[3] **J. Daniel.** *Représentation de Champs Acoustiques, Application à la Transmission et à la Reproduction de Scènes Sonores Complexes dans un Contexte Multimédia.* Université de Paris 6: Paris, France. (2000), http://gyronymo.free.fr/

[4] **C. Renard.** *Analyse objective et subjective d'une technique de rendu sonore 2D sur une zone d'écoute étendue, l'holophonie, en vue de réaliser un mur de téléprésence.* Technical Report, DEA Université du Maine (2000)

[5] **IST Project CARROUSO**: http://www.cordis.lu/ist/ka3/iaf/projects/carrouso.htm