From vibration to perception: using Large Multi-Actuator Panels (LaMAPs) to create coherent audio-visual environments

M. Rébillat[a], E. Corteel[b], B. F. Katz[c] and X. Boutillon[d]

[a]Département d'études cognitives, Ecole Normale Supérieure Paris, 29 Rue d'Ulm, 75230 Paris, France
[b]sonic emotion labs, 42 bis rue de Lourmel, 75015 Paris, France
[c]LIMSI-CNRS, BP 133 Bât 508, 91403 Orsay, France
[d]Laboratoire de Mécanique des Solides, Ecole Polytechnique, 91128 Palaiseau, France
marc.rebillat@ens.fr

Virtual reality aims at providing users with audio-visual worlds where they will behave and learn as if they were in the real world. In this context, specific acoustic transducers are needed to fulfill simultaneous spatial requirements on visual and audio rendering in order to make them coherent. Large multi-actuator panels (LaMAPs) allow for the combined construction of a projection screen and loudspeaker array, and thus allows for the coherent creation of an audio and visual virtual world. They thus constitute an attractive alternative to electro-dynamical loudspeakers and multi-actuator panels previously used. In this paper, the vibroacoustic behavior of LaMAPs is studied and it is shown that LaMAPs can be used as secondary sources for wave field synthesis (WFS). The auditory virtual environment created by LaMAPs driven by WFS is then perceptually assessed in an experiment where users estimate the egocentric distance of an audio virtual object by means of triangulation. Vibro-acoustic and perceptual results indicate that LaMAPs driven by WFS can be confidently used for the creation of auditory virtual worlds.

1 Introduction

Virtual reality aims at providing users with audio-visual worlds where they will behave and learn as if they were in the real world. To render spatialized sound in this context, holophonic technologies, such as wave field synthesis (WFS, [1]), allow for the best compromise between the size of the immersion area and the spatial rendering quality [2]. However, the integration of WFS in visual rendering systems based on large immersive screens must face practical challenges: positioning conflicts between screens and loudspeakers [3] and sound field filtering caused by projection screens [4, 5]. For virtual reality applications, specific acoustic transducers dealing with these constraints are thus needed in order to fulfill spatial requirements simultaneously on visual and audio renderings.

Multi-actuator panels (MAPs, [6, 7]) are stiff lightweight panels with multiple electro-mechanical transducers attached to the backside. Novel large dimension MAPs (LaMAPs, [2, 8]) have been designed in order to provide sufficient surface area and size to be used as a projection screen. By their design, LaMAPs avoid the positioning conflicts between screens and loudspeakers and the acoustical filtering caused by acoustic transmission loss through projection screens. LaMAPs allow for the combined creation of a screen and loudspeaker array, and thus for the coherent creation of an audio and visual virtual world. For these reasons, they constitute an attractive alternative to electro-dynamical loudspeakers and MAPs.

Sec. 2 provides a brief overview of the use of LaMAPs that was achieved in the Spatial Multi-user Audio-visual Real-Time Interactive Interface (SMART-I^2) [2, 8]. In Sec. 3, the ability of LaMAPs to provide suitable sound sources for WFS is assessed both theoretically, using a simple mechanical model, and experimentally, by measuring velocity on a LaMAP and the reconstructed sound field. Sec. 4 assesses the perceptual ability of LaMAPs to provide a convincing spatialized audio virtual environment by studying egocentric audio distance perception of static virtual sources by moving subjects.

2 LaMAPs in the SMART-I^2

This section provides a brief overview of the use of Large Multi-Actuator Panels (LaMAPs) within the Spatial Multi-user Audio-visual Real-Time Interactive Interface (SMART-I^2) [2, 8], a virtual reality rendering device based on acoustical wave field synthesis (WFS).

2.1 Wave Field Synthesis (WFS)

Wave Field Synthesis (WFS) is a spatialized sound rendering technology which was first developed at Delft University [1]. It is an audio implementation of the Rayleigh I equation, which states that: " *Every sound field emerging from one sound source can be reproduced by summing contributions of an infinite and continuous distribution of monopoles."*

Implementations of WFS are simplified versions of this principle often using a horizontal linear array of equally spaced loudspeakers [5]. As auditory perception is more precise and more stable in the horizontal plane [9], this is a pertinent choice for array orientation. This allows for a reduction in the required computational power.

If an array of monopoles is available, virtual point sources can be synthesized at positions behind or in front of the array. Due to the discrete spatial distribution of loudspeakers, the sound field corresponding to these point sources is physically correctly reconstructed up to an aliasing frequency f_{al} which depends on loudspeaker spacing. Focused sources (*i.e.* point sources in front of the array) are perceived as being physically present in the immersion area.

2.2 Large Multi-Actuator Panels (LaMAPs)

The integration of WFS with a 3D visual rendering technology is achieved through an innovative use of multi-actuator panels (MAPs) [6, 7]. MAPs are stiff lightweight panels with multiple electro-mechanical transducers attached to the backside, typically not larger than 1 m^2. For this project, a novel large dimension MAP (LaMAP) has been designed (*i.e.* 5 m^2 with a 4/3 ratio) in order to provide sufficient surface area and size to be used as a projection screen. Such a structure then allows one to efficiently integrate a 3D visual rendering technology with WFS.

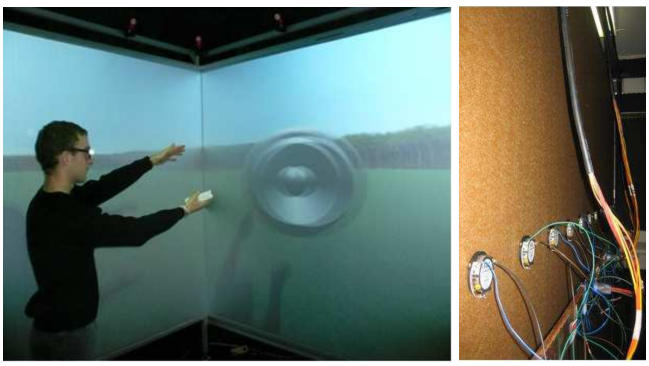

Figure 1: *Left:* Example of AV scene rendered by the SMART-I^2. *Right:* Back face of the SMART-I^2.

In the SMART-I^2 system, two LaMAPs of 2.6 m × 2 m form a corner of stereoscopic screens and a 24 channel loudspeaker array. With this configuration, users can move within an immersion area of approximately 2.5 m × 2.5 m. Electrodynamic exciters are spaced by 21 cm thus allowing for a physically correct synthesis of the sound field up to $f_{al} \simeq$ 1.5 kHz. The frequency response of the LaMAPs is approximately flat between 100 Hz and 16 kHz. Fig. 1 provides an example of a user in an AV scene rendered by the SMART-I^2 and shows the transducers attached to the back on one LaMAP.

3 Vibroacoustic behavior of LaMAPs

In this section, the ability of LaMAPs to provide suitable sound sources for WFS, *i.e.* monopoles, up to the aliasing frequency is assessed both theoretically, using a simple mechanical model, and experimentally, by measuring velocity on a LaMAP and the reconstructed sound field.

3.1 A simple mechanical model of the LaMAPs

As a first approximation, LaMAPs have been modelled as infinite, homogeneous, isotropic, thin plates [10]. LaMAPs are characterized by their complex rigidity $D = D_0(1 + j\eta)$ with $D_0 = 7.5$ kg×m^2/s^2 and $\eta = 0.01$, and their mass density $\mu = 424$ g/m^2 [11]. The dispersion equation, relating the wavenumber k to the angular frequency ω in LaMAPs, is:

$$Dk^4 = \mu\omega^2 \tag{1}$$

The coincidence frequency f_c, which characterises the frequency below which the panel excited ponctually radiates almost omnidirectionally [10], can be computed according to this simple model:

$$f_c = \frac{c^2\mu^{1/2}}{2\pi D_0^{1/2}} \tag{2}$$

with $c = 340$ m/s representing the speed of sound in air. The coincidence frequency of the panels of the SMART-I^2 is $f_c \simeq 4.3$ kHz.

The contact zone between exciters and LaMAPs is a nearly ring with a diameter $d = 3.5$ cm. According to Eq. (1), the wavelength $\lambda = 2\pi/k$ is greater than d for frequencies lower than 2.1 kHz. Exciters can thus be supposed to act as point sources on LaMAPs below this frequency.

Therefore, one may safely consider that LaMAPS act as monopoles (punctual and omnidirectional sources) below the aliasing frequency $f_{al} = 1.5$ kHz described in Sec. 2.1.

3.2 Velocity response near one exciter

The velocity response of an infinite point-excited LaMAP is given by [10]:

$$V(k, r) = V_0 \left[H_0^{(2)}(kr) - H_0^{(2)}(-jkr) \right] \tag{3}$$

with $j^2 = -1$, k the wavenumber given by Eq. (1) at ω, r the distance to the excitation point, and $H_0^{(2)}(.)$ the Hankel function of second kind of order 0.

To investigate the validity of this simple model, velocity impulse responses (obtained by means of a LASER vibrometer and using the method presented in [12]) have been measured along a 90 cm vertical line of 101 points, symmetrically distributed above and below one exciter located in the middle of the right LaMAP (1.2 m from the right boundary and 0.9 m from the top).

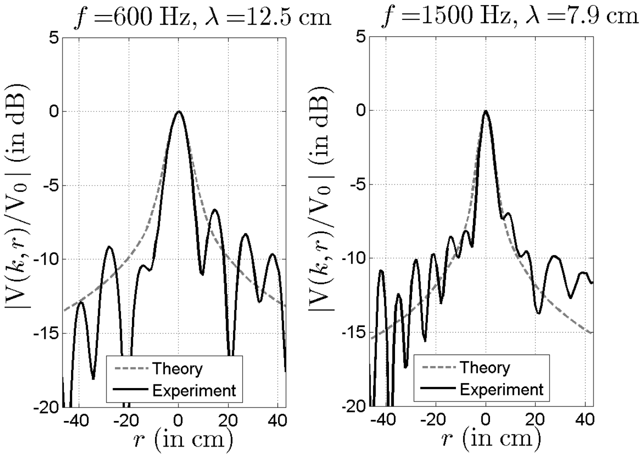

Figure 2: Experimental and theoretical (infinite point-excited LaMAPs) velocity profiles.

Theoretical and experimental velocity profiles are presented in Fig. 2 for $f = 600$ Hz and $f = 1.5$ kHz. Far from the excitation point, experimental profiles exhibit notches and peaks at positions related to wavelengths. These are ignored by the theory of infinite point-excited LaMAPs. Boundary effects cannot be completely ignored far from the excitation point. Near the excitation point, theoretical predictions are in very good agreement with experimental data.

These measurements thus validate the hypothesis that exciters act as point sources on LaMAPs for frequencies below the aliasing frequency.

3.3 Sound field reconstructed using LaMAPs

According to Secs. 3.1 and 3.2, LaMAPs can be considered as approximate monopole sources below the aliasing frequency $f_{al} \simeq 1.5$ kHz. They can thus potentially synthesize physically correct sound fields when driven by WFS.

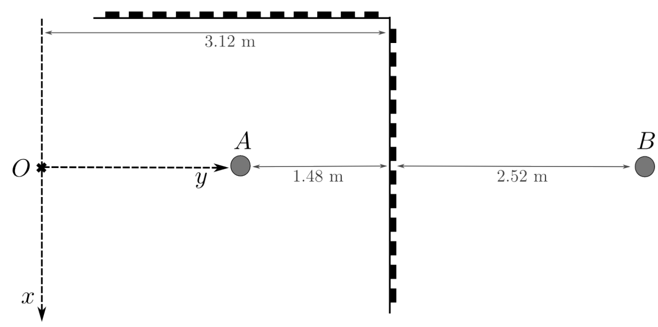

Figure 3: Experimental protocol used for the measurement of acoustical pressure impulse responses.

To validate this point, acoustical pressure fields generated by LaMAPs driven by WFS have been measured. Acoustical pressure impulse responses have been acquired on 30 points spaced by 9 cm located on a x-line parallel to the right LaMAP and situated at 3.12 m from it (see Fig. 3). Impulse responses have then been low-pass filtered to retain only frequencies below $f_{al} = 1.5$ kHz. Low-pass filtered acoustical pressure impulse responses are represented in the spatio-temporal domain for a point source A located 1.48 m in front of the right LaMAP (focused source) in Fig. 4 and for a point source B located 2.52 m behind the right LaMAP in Fig. 5.

Theoretically, an acoustical pressure impulse emitted at $t = 0$ ms by the source A (or B) should be received on the x-line at a time $t_r(x)$ such that:

$$t_r(x) = \frac{1}{c}\sqrt{y_{A/B}^2 + x^2} + t_0 \quad (4)$$

This constitutes the ideal curve that should be observed in the spatio-temporal domain.

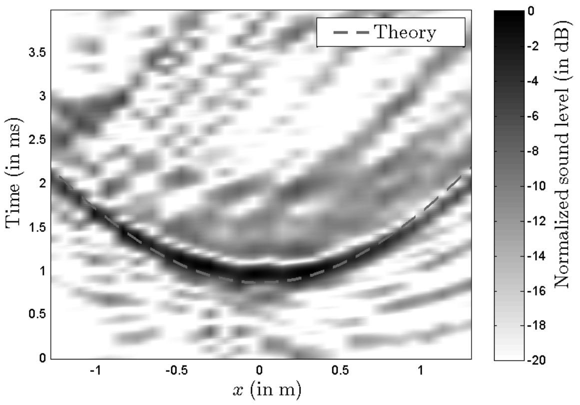

Figure 4: Spatio-temporal domain representation of the low-pass filtered acoustical pressure impulse responses for a point source A located 1.48 m in front of the right LaMAP.

From Fig. 4, it can be observed that, for a virtual source in front of the LaMAP, contributions of each exciter interfere in a constructive manner to synthesize the corresponding wave front. As expected, individual contributions of each exciter can be seen as arriving as "pre-echoes" with a lower level before the wave front to be synthesized. Reflections coming from the left LaMAP are visible. The agreement between theory and experiment is furthermore very good.

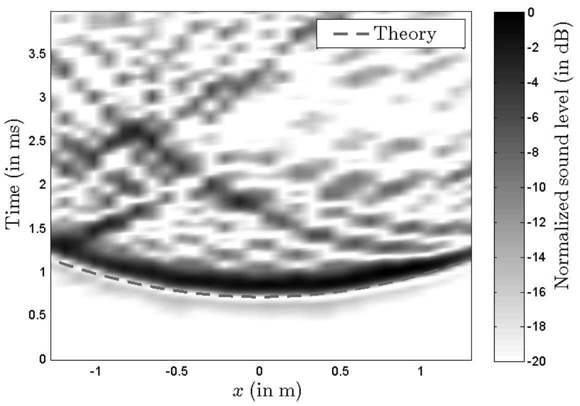

Figure 5: Spatio-temporal domain representation of the low-pass filtered acoustical pressure impulse responses for a point source B located 2.52 m behind the right LaMAP.

From Fig. 5, it can be seen that for a virtual point source situated behind the right panel contributions of each exciters also interfere in a constructive manner in order to synthesize the corresponding wave front. After the first wave front, individual contributions of each exciters can be seen, as well as reflections coming from the left LaMAP. The agreement between theory and experiment is still very good.

These measurements thus validate from a vibro-acoustical perspective the use of LaMAPs as monopole sound sources for WFS up to the aliasing frequency and show that they allow to synthesize physically correct sound fields when driven by WFS.

4 Distance perception using LaMAPs

The previous section has validated the ability of LaMAPs to physically synthesize wave fields corresponding to virtual sound sources in the SMART-I^2. This section now assesses perceptually the ability of LaMAPs to provide a convicing spatialized audio virtual environment by studying egocentric audio distance perception of static virtual sources by moving subjects [2, 13].

4.1 Experimental design

A total of 20 volunteers between 24 and 45 years old participated in the experiment (17 men). All subjects had self-reported normal vision (possibly corrected) and normal hearing. Five distances were tested: 1.5 m, 2 m, 2.7 m, 3.5 m, 5 m (the LaMAP was located at a distance of 2.3 m). Each subject had to estimate the five distances four times in a random order.

The audio target object was a 4 kHz low-pass filtered white noise with a 15 Hz amplitude modulation. Low-pass filtered white noise has been chosen in order to have a wide spectral content and to allow subjects to rely on numerous audio localization cues. The white noise was modulated in amplitude by a sine wave to produce attack transients that are also useful in sound localization [9]. No simulated room-effect (*i.e.* ground reflections) was included. The sound level of the audio object corresponds to a monopole emitting 78 dB(SPL) at 1 m.

Audio target objects were presented together with a realistic audio-visual environment. The visual environment consisted of an open, grassy field with a forest in the background (see Fig. 1). The associated audio environment consisted of the sound of wind in the trees accompanied by some distant bird songs (overall background level of 36 dBA). The audio environment was spatialized using 12 plane waves equally distributed in the horizontal frontal field of rendering. Environmental sound levels were adjusted to be slightly above the background noise produced by the four video-projectors (background noise level of 34 dBA).

4.2 Presentation and report

Distance estimation was performed here in two phases: a *presentation* phase, and a *reporting* phase (see Fig. 6). Subjects began each iteration at the *start position*, indicated by a black "×" in Fig. 6. In the *presentation* phase, subjects move around in the *exploration* area which was a rectangle of 1×0.8 m^2. During this phase, subjects were instructed to acquire "*a good mental representation of the virtual object and of its environment*". Before starting this phase, subjects had to indicate that they were ready to perform the *presentation* phase (pressing a wiimote button).

Once "*a good mental representation*" has been acquired, subjects were asked to press a wiimote button to indicate that

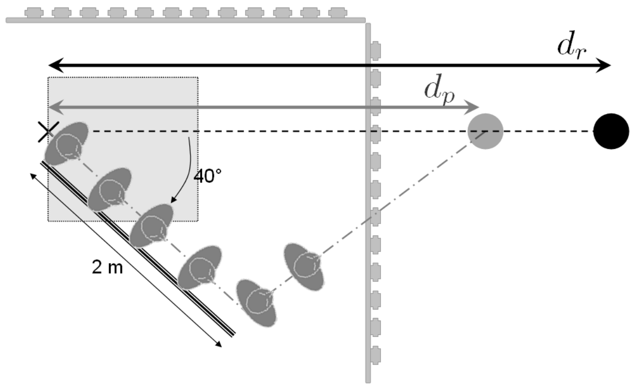

Figure 6: Presentation and reporting phases.

they were ready for the *reporting* phase. At this point, the perceived distance d_p was estimated by means of triangulated blind walking [14, 15]. Subjects closed their eyes, made a 40° right-turn and walked blindly for an imposed distance of $\simeq$ 2 m, following a handrail guide which was present to help during blind-walking. Subjects stopped at the end of the guide, turned in the direction where they thought the object was, and took a step forward in the direction of the source position. They then indicated that they had completed the reporting by pressing a wiimote button. Afterwards, they could open their eyes and go back to the *start position* for the next iteration. The experimental protocol was fully automated, with the subjects being observed remotely so as not to disturb the sense of presence.

4.3 Results

The position of the head of the subject (central point between the eyes) and corresponding times are recorded for each iteration during both the *presentation* and *reporting* phases at 100 Hz. Perceived distances d_p were estimated from the *triangulation* trajectory as follows. A line ($y = ax + b$) was fitted to the trajectory points during the forward step. The estimated perceived distance is given by Eq. (5):

$$d_p = -\frac{b}{a} \tag{5}$$

Mean and standard deviation of the perceived distance d_p as a function of rendered distance d_r are presented in Fig. 7.

4.4 Discussion

A compressive curve in the form of $d_p = k(d_r)^a$ has been shown to be a good model for the psychophysical function that relates estimates of perceived distance to physical source distance [16]. A review among 84 data sets presented in [16] reported mean values of $k = 1.32$ and $a = 0.54$ when fitting a compressive model to these data of previous studies with real acoustic sources. Such a compressive model is thus an interesting psychophysical curve to compare with the data obtained here with virtual sources.

Results are compared to the compressive model in Fig. 7. By fitting such a model to the perceived audio distances collected in the present study, values of $a = 0.34 \pm 0.085$ and $k = 1.69 \pm 0.31$ m are found, with $R^2 = 98\%$ of the variance observed in the experimental data explained by the compressive model. The model $d_p = k(d_r)^a$ thus fits very well to the experimental data. The perception of auditory distance

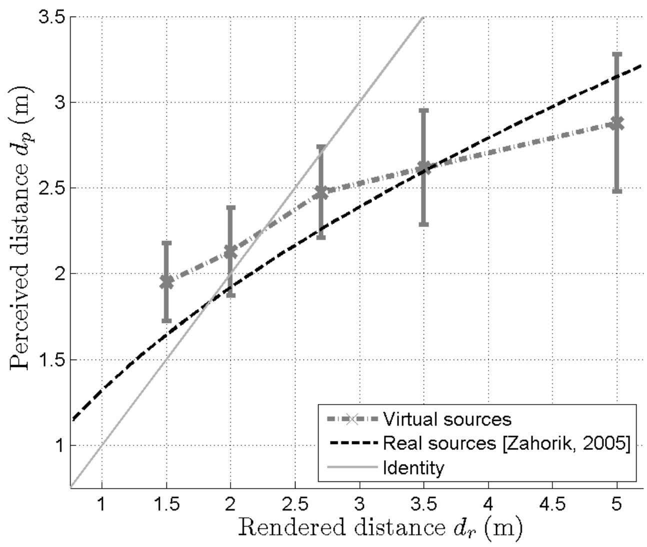

Figure 7: Mean and standard deviation of the perceived distance d_p as a function of rendered distance d_r. Vertical bars indicate one standard deviation.

seems to be slightly more compressed in the virtual world than predicted in the real world using the average compressive model.

It can thus be concluded that LaMAPs driven by WFS are able to synthesize sound-fields which are perceptually meaningful in terms of distance for moving subjects and for static virtual sources between 1.5 m and 5 m, but with apparently slightly more compression than in the real world.

5 Conclusion

In the context of virtual reality, specific acoustic transducers are needed to fulfill spatial requirements imposed simultaneously on visual and audio rendering in order to make them coherent. Large multi-actuator panels (LaMAPs) allow for the construction of a combined projection screen and loudspeaker array, enabling the coherent creation of an audio and visual virtual world. They thus constitute an attractive alternative to electro-dynamical loudspeakers and multi-actuator panels, used so far. In the SMART-I^2 system, two LaMAPs form a corner of stereoscopic screens and a loudspeakers array, allowing users to move within a large immersion area. Using wave field synthesis (WFS), a physically correct synthesis of the sound fields up to $f_{al} \simeq 1.5$ kHz is expected here.

From a vibro-acoustical perspective, it is predicted theoretically that LaMAPs approximately act as monopoles below the aliasing frequency f_{al} of the SMART-I^2. Mechanical velocity measurements validate the hypothesis that exciters act as point sources on LaMAPs below f_{al} and acoustic pressure field measurements show that LaMAPs allow one to physically synthesize correct sound fields when driven by WFS. From a perceptual perspective, it is shown that LaMAPs driven by WFS are able to synthesize sound-fields which are perceptually meaningful in terms of distance for moving subjects and for static virtual sources between 1.5 m and 5 m, but with slightly more compression than has been shown for real world sources.

Vibro-acoustic and perceptual results thus both indicate that LaMAPs driven by WFS can be confidently used for the creation of *auditory* virtual worlds. Additional experimental results (not presented here, see [2, 13]) have furthermore shown that this auditory virtual world is perceived as spatially coherent with the 3D-visual world provided by the SMART-I^2. LaMAPs driven by WFS can thus be safely used for the creation of spatially coherent *auditory-visual* virtual worlds.

Acknowledgements

All the work presented here was performed during the first author's Ph.D. thesis at the LIMSI (CNRS) and LMS (CNRS and École Polytechnique). LaMAPs as well as the audio amplifiers and the wave field synthesis sound rendering engine have been kindly provided by *sonic emotion labs*. Authors would like to thank Aurélien Renoncourt (internship L3 Physique Mécanique, Université Paris-Sud) for performing the sound field measurements presented in Sec. 3.3. Thomas Chartier and Philippe Cuvillier (internship, École Polytechnique, Palaiseau), are also thanked for their help in the preliminary tests of the experiment presented in Sec. 4.

References

[1] A. J. Berkhout, D. de Vries, and P. Vogel. Acoustic control by wave field synthesis. *Journal of the Acoustical Society of America*, 93(5):2764–2778, 1993.

[2] M. Rébillat. *Vibrations de plaques multi-excitateurs de grandes dimensions pour la création d'environnements virtuels audio-visuels: approches acoustique, mécanique, et perceptive.* PhD thesis, École Polytechnique, 2011.

[3] D. de Vries. *Wave Field Synthesis*. Audio Engineering society, 2009.

[4] J. Hiipakka, T. Ilmonen, T. Lokki, and L. Savioja. Sound signal processing for a virtual room. In *EUPSICO 2000 : European Signal Processing Conference No 10, Tampere , Finlande*, 2000.

[5] E. Corteel. Equalization in an extended area using multichannel inversion and wave field synthesis. *Journal of the Audio Engineering Society*, 54(12):1140–1161, 2006.

[6] M. M. Boone. Multi-actuator panels (MAPs) as loudspeaker arrays for wave field synthesis. *Journal of the Audio Engineering Society*, 52(7-8):712–723, 2004.

[7] B. Pueo, J. J. Lopez, J. Escolano, and Lars Horchens. Multiactuator panels for wave field synthesis: Evolution and present developments. *Journal of the Audio Engineering Society*, 58(12):1045–1063, 2010.

[8] M. Rébillat, E. Corteel, and B.F.G. Katz. SMART-I^2: "Spatial Multi-user Audio-visual Real-Time Interactive Interface". *125th Convention of the Audio Engineering Society*, 2008.

[9] J. Blauert. *Spatial Hearing, The Psychophysics of Human Sound Localization*. MIT Press, 1999.

[10] L. Cremer, M. Heckl, and B.A.T. Peterson. *Structure-Borne Sound: Structural Vibrations and Sound Radiation at Audio Frequencies*. Springer-Verlag, 2005.

[11] M. Rébillat and X. Boutillon. Measurement of relevant elastic and damping material properties in sandwich thick-plates. *Journal of Sound and Vibration*, 330(25):6098–6121, 2011.

[12] M. Rébillat, R. Hennequin, E. Corteel, and B. F.G. Katz. Identification of cascade of Hammerstein models for the description of nonlinearities in vibrating devices. *Journal of Sound and Vibration*, 330:1018–1038, 2011.

[13] M. Rébillat, X. Boutillon, E. Corteel, and B. F. G. Katz. Audio, visual and audio-visual egocentric distance perception in virtual environments. In *Proceedings of the Forum Acusticum, Aalborg, Danemark*, 2011.

[14] S. S. Fukusima, J. M. Loomis, and J. A. Da Silva. Visual perception of egocentric distance as assessed by triangulation. *Journal of Experimental Psychology-Human Perception and Performance*, 23(1):86–100, February 1997.

[15] J. M. Loomis, R. L. Klatzky, J. W. Philbeck, and R. G. Golledge. Assessing auditory distance perception using perceptually directed action. *Perception & Psychophysics*, 60(6):966–980, 1998.

[16] P. Zahorik, D. S. Brungart, and A. W. Bronkhorst. Auditory distance perception in humans: A summary of past and present research. *Acta Acustica United with Acustica*, 91(3):409–420, May 2005.