Objective and subjective comparison of electrodynamic and MAP loudspeakers for Wave Field Synthesis

Etienne Corteel[1,2], Khoa-Van NGuyen[1], Olivier Warusfel[1], Terence Caulkins[1], Renato Pellegrini[2]

[1]*IRCAM, 1. Pl. Igor Stravinsky, 75004 Paris, France*

[2]*sonic emotion, Eichweg 4, CH-8154 Oberglatt, Switzerland*

Correspondence should be addressed to Etienne Corteel (`etienne.corteel@sonicemotion.com`, `etienne.corteel@ircam.fr`)

ABSTRACT

This article deals with the direct comparison of WFS rendering using either MAP or electrodynamic loudspeakers on an objective and a subjective level. Objective criteria are used to evaluate coloration and localisation cues that are perceived in an extended listening area. It is shown in a first listening test that the reduced spatial coherence of MAP loudspeakers partly explains the perceived differences between the two loudspeaker technologies. This "diffuse" behavior can be artificially produced on electrodynamic loudspeakers using a diffuse filtering. The proposed diffuse filtering may also limit rendering artifacts above the spatial aliasing frequency. Finally, it is shown in a second listening experiment that MAP loudspeakers favor distance perception compared to electrodynamic loudspeakers.

0. INTRODUCTION

Wave Field Synthesis (WFS) is a multichannel sound rendering technique that allows for the synthesis of physical properties of sound fields within an extended listening area [1]. It relies on a large number of closely spaced (typically 15-20 *cm*) loudspeakers (typically 15-20 *cm*) loudspeakers forming one or several linear an acoustic aperture through which the target sound field (as emanating from a target sound source) propagates into the listening environment.

Practical implementation of WFS requires simplifications to the underlying physical principles (Kirchhoff-Helmholtz and Rayleigh integrals). Real loudspeakers radiation characteristics may also contribute to alter the synthesized sound field compared to the target one. The perceptual impact of these inaccuracies relates to the more general problem of transparency of the sound rendering medium. Ideally, this medium should not be detected anywhere inside the listening area so as to create an illusion of non-mediation [2]. In this paper, we consider an extended definition of transparency that does not only include timbre but also spatial aspects such as localization cues (angular position, distance) as well as spatial impression (room-size, reverberance, ...). Moreover, transparency evaluation should account for non-acoustic cues [3] such as the feeling of wearing headphones or seeing loudspeakers.

Two types of loudspeakers are used nowadays for Wave Field Synthesis (see figure 1):

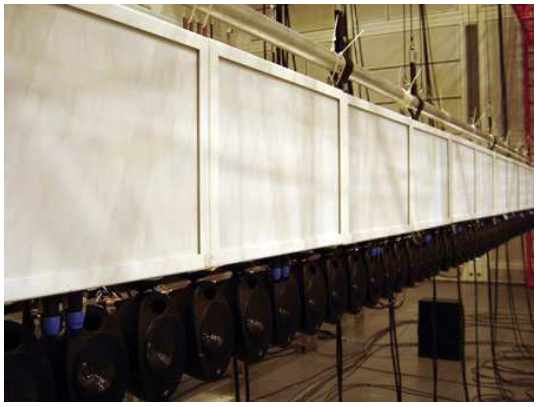


Fig. 1: MAP and electrodynamic loudspeakers

- array-mounted electrodynamic loudspeakers,

- Multi-Actuator Panels (MAP).

MAP loudspeakers have been recently proposed as an alternative to electrodynamic loudspeakers for WFS [4]. Thanks to their low visual profile, MAP loudspeakers were originally thought as a way to facilitate the integration of tens to hundreds of loudspeakers in an existing environment. Informal listening tests with researchers and composers, made at IRCAM and at the Forum Neues Musiktheater in 2005, indicated that MAP loudspeakers may improve transparency, especially in terms of distance perception. It was thus decided to conduct systematic objective and subjective studies to compare the acoustic properties of both loudspeaker types for WFS rendering.

The goal of this paper is to compare, at an objective and a subjective level, the transparency of Wave Field Synthesis rendering using electrodynamic or MAP loudspeakers. The influence of non-acoustic factors such as visual cues are beyond the scope of this paper.

In a first section, the radiation properties (directivity, spatial coherence [5]) of both electrodynamic and MAP loudspeakers are shown. Since MAP loudspeakers rely on Distributed Mode Loudspeaker (DML) technology, they exhibit a "diffuse" behavior (reduced spatial coherence), especially at high frequency. In a second section, diffuse filtering for WFS is introduced. It is meant as a way to replicate the diffuse properties of MAP on electrodynamic loudspeakers so as to validate their potential benefit on transparency. In a third section, simple objective criteria are proposed. They account for acoustic dimensions (coloration, localization cues) that may contribute to the transparency of the sound reproduction system within an extended listening area. This objective analysis is finally completed with two subjective listening experiments: on the discrimination of loudspeaker and filtering types in an ABX test, and on the perceived distance in a pair-comparison test.

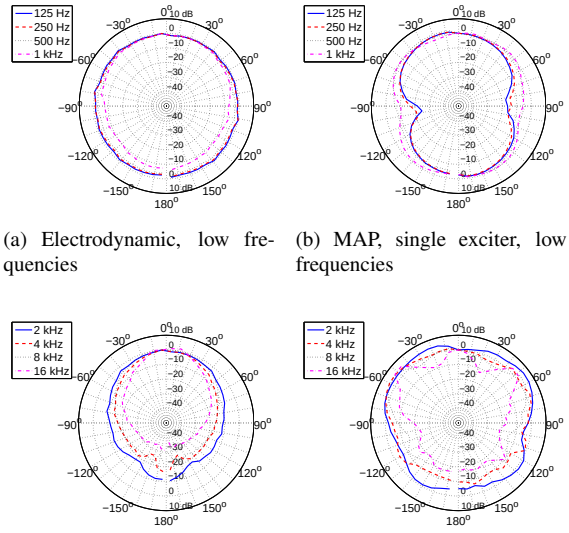
1. LOUDSPEAKERS FOR WAVE FIELD SYNTHESIS

In this section, the radiation properties of both types of loudspeaker are described and compared considering a single transducer. The electrodynamic loudspeaker is manufactured by Kef, model KHT 2005. The MAP loudspeaker is manufactured by sonic emotion.

1.1. Directivity

Figures 2(a), 2(b), 2(c), and 2(d) display octave band directivity polar plots of both electrodynamic and MAP loudspeakers. Measurements were achieved in an anechoic chamber using 48 regularly spaced microphone positions (7.5° precision) at 1.4 *m* distance from the loudspeakers. In these plots, the levels are normalized in reference to the frontal direction (0°) so as to display only directivity information. Figures 2(a) and 2(b) display polar plots at "low frequencies" (125 Hz, 250 Hz, 500 Hz, 1000 Hz). Figures 2(c) and 2(d) display polar plots at "high frequencies" (2 kHz, 4 kHz, 8 kHz, 16 kHz).

MAP loudspeakers have more complex directivity properties at low frequencies compared to electrodynamic loudspeakers. This can be problematic for WFS since the theory assumes that loudspeakers have ideal omnidirectional behavior. Multichannel equalization may be used to compensate for such non ideal loudspeakers directivity [6] [7]. This method also reduces artifacts inherent to WFS at low frequencies (near-field artifacts, diffraction). At higher frequencies electrodynamic loudspeakers become more and more directive whereas MAP loudspeakers exhibit an irregular although more "omnidirectional" behavior in average.



(a) Electrodynamic, low frequencies

(b) MAP, single exciter, low frequencies

(c) Electrodynamic, high frequencies

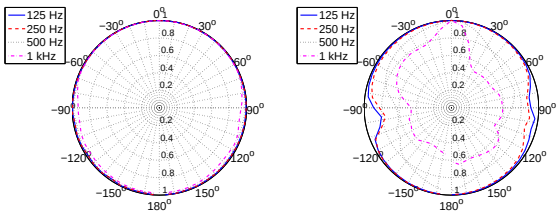
(d) MAP, single exciter, high frequencies

Fig. 2: Loudspeaker radiation pattern

1.2. Spatial coherence

Figures 3(a), 3(b), 3(c), and 3(d) display octave band Cross-Correlation Function (CCF) polar plots of both electrodynamic and MAP loudspeakers. CCF polar plots were introduced by Gontcharov and Hill in [5] to de-

scribe diffuse radiation properties of Distributed Mode Loudspeakers (DML). They are obtained by extracting the maximum of the cross-correlation function calculated from the pair of band-pass filtered impulse responses measured in the considered and the frontal (0°) direction. Such a quantity thus provides an estimate of the "spatial correlation" of the loudspeaker radiation.
Figures 3(a) and 3(b) display polar plots at "low frequencies" (125 Hz, 250 Hz, 500 Hz, 1000 Hz). Figures 3(c) and 3(d) display polar plots at "high frequencies" (2 kHz, 4 kHz, 8 kHz, 16 kHz).
Both MAP and electrodynamic loudspeakers have high spatial correlation at 125 Hz and 250 Hz. For MAP loudspeakers, CCF values reduce above 500 Hz due to the more complex directivity characteristics of these loudspeakers and their "diffuse" properties. This is potentially detrimental for WFS which relies on phase interference of coherent wave fronts emitted by loudspeakers.



(a) Electrodynamic, low frequencies

(b) MAP, single exciter, low frequencies

(c) Electrodynamic, high frequencies

(d) MAP, single exciter, high frequencies

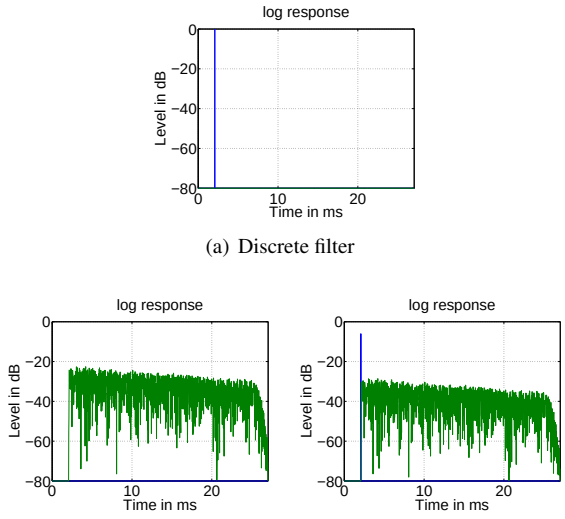
Fig. 3: Loudspeaker spatial correlation pattern

2. DIFFUSE FILTERING AND ITS CONSEQUENCES ON THE SPATIAL RESPONSE OF THE LOUDSPEAKER ARRAY

In this section, diffuse filtering for WFS rendering is first introduced. The impact of diffuse filtering is then evaluated by computing the spatial response of the loudspeaker array. In this section, ideal omnidirectional loudspeakers are used.

2.1. Diffuse filtering

Figures 4(a), 4(b) and 4(c) display impulse responses of filters that may be used for WFS rendering. The impulse response of figure 4(a) is a "classical" WFS filter (a delayed, possibly attenuated dirac pulse).
Figure 4(b) displays the impulse response of a "diffuse" filter that will be referred to as "Full Diffuse" (FD). This filter is generated from a time limited white noise that is generated independently for each loudspeakers in order to obtain uncorrelated outputs. The temporal envelope of the noise is modified by applying a window with a sharp attack and a decay slope. The short length of the noise and the modification of its temporal structure may introduce coloration artifacts. A whitening process is thus applied. It consists in adjusting the level in auditory (ERB_N) frequency bands [8].
A third type of filter is displayed in figure 4(c). It appears as a combination of the "discrete" and the "diffuse" filter and is therefore referred to as Discrete-Diffuse (DD) filter. This filter is designed in such a way that half of the energy is provided by the discrete part, the other half by the diffuse part. This defines a more general class of diffuse filtering for which the total energy is divided into the discrete and the diffuse part.



(a) Discrete filter

(b) Diffuse filter (FD, Full Diffuse)

(c) Discrete-diffuse filter (DD)

Fig. 4: Various diffusing filters used at high frequencies

2.2. Spatial response

The spatial response of a loudspeaker array may be ob-

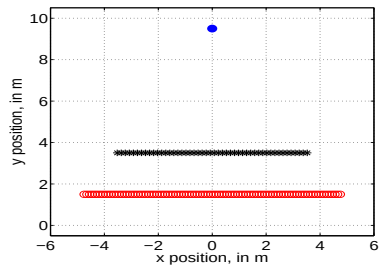


Fig. 5: Test configuration, loudspeakers (black *), microphones (red o), virtual source (blue .), top view.

tained by first measuring (simulating) each channel on a linear microphone array parallel to the loudspeaker array. The measured (or simulated) impulse responses are then convolved with the calculated filters for the synthesis of a given source. The response of the loudspeaker array is finally obtained by summing the contributions of all loudspeakers at each measurement position.

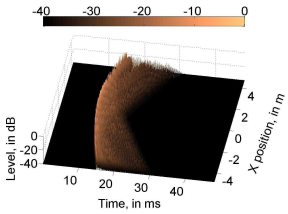
We consider here a linear array comprising 48 ideal omnidirectional loudspeakers with 15 *cm* spacing (see figure 5). The microphone array is composed of 96 ideal omnidirectional microphones with 10 *cm* spacing. It is situated 2 *m* from the loudspeaker array. The target virtual source is centered and located 6 m behind the loudspeaker array (see figure 5). Filters are calculated at low frequencies using the previously mentioned multichannel equalization method [7].

One of the main limitations of Wave Field Synthesis is known as spatial aliasing. Spatial aliasing is due to the spatial sampling of the loudspeaker distribution so as to limit the number of channels [9]. The corresponding Nyquist frequency is referred to as the spatial aliasing frequency. Below the spatial aliasing frequency, contributions from all loudspeakers of the loudspeaker array fuse into a single target wave front. This is not the case above the aliasing frequency and the sound field cannot be controlled in an extended listening area.

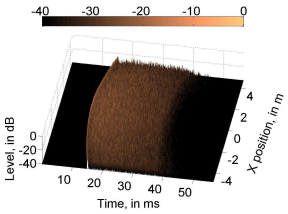
Above the aliasing frequency, the three type of filters are used (Discrete, Full-Diffuse FD, Discrete-Diffuse DD). The combination of multichannel equalization at low frequencies and discrete filter above the aliasing frequency is normally used for multichannel equalization [7]. This filtering process will therefore be referred to as "MEQ".

2.2.1. Impulse response

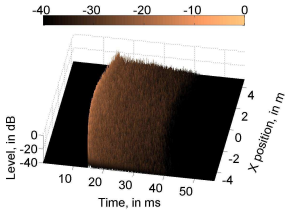
Figures 6(a), 6(b), and 6(c) display spatial temporal response of the loudspeaker array. It can be seen that it



(a) Discrete filter



(b) Diffuse filter (FD, Full Diffuse)



(c) Discrete-diffuse filter (DD)

Fig. 6: Impulse responses, configuration displayed in figure 5, dependency on diffuse filter characteristics

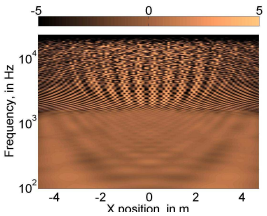
significantly differs from a delayed and attenuated dirac pulse (ideal response). Above the spatial aliasing frequency, all loudspeaker contributions create a temporally spread impulse response. However, it can be shown that the first wave front (attack of the response) remains consistent with the expected propagation time given the distance between the virtual source and the measurement position.

For the MEQ filter, the length of the non-null temporal response (effective length) varies from 10 ms at center positions to almost 20 ms to the sides (see figure 6(a)). This is related to the finite length of the loudspeaker array [10] [6].

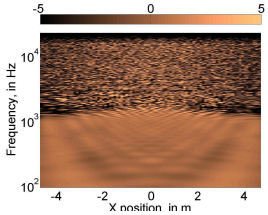
For DD and FD filters the temporal response is longer (about 25 ms along asides) because of the temporally spread response of the filters (see figures 6(b) and 6(c)). However, the attack of the DD filter response attack appears somewhat sharper than that of the FD filter response because of its "discrete" component.

2.2.2. Frequency response

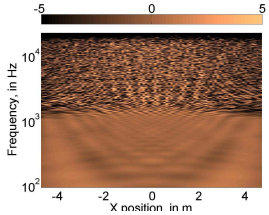
Figures 7(a), 7(b), and 7(c) display the spatial frequency response of the loudspeaker array for the three types of filtering. The influence of the absolute level at each position has been removed by the applying a normalization factor. These responses therefore only display the deviation from the target level (0 dB).



(a) Discrete filter



(b) Diffuse filter (FD, Full Diffuse)



(c) Diffuse-discrete filter (DD)

Fig. 7: Frequency responses, level in dB, color scale indicated in the color bar on top of each figure, configuration displayed in figure 5, dependency on diffuse filter characteristics

It can be seen that, at low frequencies, the response is almost flat at all microphone positions with only small oscillations that remain within $\pm$ 0.5 dB. Above the aliasing frequency spatial responses are more disturbed. They exhibit a "fast varying" response along both frequency and space axis. Above the aliasing frequency, interference patterns can easily be noticed for the MEQ filter (see figure 7(a)). These may be detected with listener's movements within the listening area [11]. These patterns do not appear with DD and FD filters (see figures 7(b) and 7(c)) since the diffuse filtering limits correlation of the loudspeakers' contributions to the synthesized sound field. In the case of the DD filter, only half of the energy is "diffuse" but this seems to be sufficient to avoid visually noticeable patterns in the spatial frequency response. This could be verified in informal listening tests (cf. section 3.3.3).

3. OBJECTIVE COMPARISON

In this section, we propose an evaluation of the rendering system transparency focusing on two acoustic dimensions: timbre and angular localisation. The evaluation relies on objective criteria calculated from free-field spatial response of loudspeaker arrays. These objective criteria are derived from psychoacoustical experiments and physiological studies.

3.1. Test setup

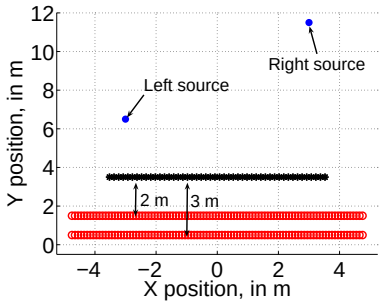


Fig. 8: Test setup: loudspeakers (black *), microphones (red 0), source (blue .), top view

The loudspeaker array layout is identical to the 48 channel system described in the previous section. Electrodynamic or MAP loudspeakers are compared using the same configuration. Both "real" arrays are measured in a large room sufficiently far away from any reflecting surface so as to window out any reflected contributions and therefore extract the free-field radiation of each loudspeaker. A 24 each loudspeaker. A 24 channel, 2.4 *m* long, microphone array is spanning 9.6 m along 2 lines parallel to the loudspeaker array at 2 and 3 m distance (see figure 8).

Two test sources are studied in this experiment (figure 8). 8). The "left source" is positioned 3 *m* to the left loudspeaker array. The "right source" is positioned 3 m to the right and 8 m behind the loudspeaker array. Filters are designed using the multichannel equalization method described in [6] and [7] based on the free-field radiation measurements at 2 *m*. Diffuse energy is introduced in the filters only above the aliasing frequency.

The spatial response of the loudspeaker arrays is then estimated from the calculated filters convolved with the free field measurements at 3 *m* distance. As a comparison, we will also consider simulated responses of an identical array composed of ideal omnidirectional loudspeakers at the same distance.

Given the test sources, the loudspeaker array geometry and the listening positions, one may expect an average spatial aliasing frequency of about 1200 Hz. This corner frequency is used throughout the objective comparison as a limit between "low" frequency and "high" frequency analysis.

3.2. Estimation model

We propose to evaluate and compare the impulse response of the system $h_\Psi(x,y,t)$ ($H_\Psi(x,y,f)$ in the frequency domain) with an "ideal" WFS response $a_\Psi(x,y,t)$ ($A_\Psi(x,y,f)$ in the frequency domain) corresponding to an infinite continuous linear array (i.e. no aliasing, no diffraction) which is expressed as:

$$a(x,y,t) = \sqrt{\frac{d_M^L + d_\Psi^L}{d_M^L}} \frac{1}{d_\Psi^M} \delta\left(t - \frac{d_\Psi^M}{c} + \tau_{eq}\right), \quad (1)$$

where d_M^L is the distance between the listening position (x, y) and the loudspeaker array, d_Ψ^L the distance of the virtual source Ψ to the loudspeaker array, and d_Ψ^M the distance between the virtual source and the listening position. This expression accounts for the propagation time of waves emitted by the virtual source and the modified attenuation law due to the use of a linear loudspeaker array for WFS. The same formula defines the target sound field in the multichannel equalization method [7]. The y dependency will be omitted for clarity since all responses are evaluated at 3 m from the loudspeaker array.

A direct comparison between the cues associated to the ideal ($a(x,t)$) and synthesized ($h(x,t)$) sound field may allow to determine acoustic factors that influence the transparency of the sound reproduction medium.
Alternatively, a quality function $q_\Psi(x,t)$ ($Q_\Psi(x,f)$ in the frequency domain) may be defined as:

$$Q_\Psi(x,f) = \frac{H_\Psi(x,f)}{A_\Psi(x,f)} \quad (2)$$

Ideally, the time domain quality function should be a dirac (constant level of 1 in the frequency domain). It incorporates all deviations (time and frequency based) between the ideal and synthesized response, and gets rid of absolute level and propagation time.

3.3. Coloration

The coloration analysis employs criteria derived from objective/subjective studies based on deviations of the frequency response. We study both the coloration which may be perceived at a fixed position and the spatial color variation while wandering in the sound installation.

3.3.1. Coloration at a fixed position

An objective criterion of coloration was recently proposed by Moore and Tan in [12]. This criterion estimates coloration as the variation of excitation level (energy in frequency bands) across $\mathrm{ERB_N}$ bands in an impulse response measured at a given position. For the sake of simplicity, we extract excitation levels in $\mathrm{ERB_N}$ band i from the quality function $Q_\Psi(x,f)$ at position x as:

$$EQ_i(x) = 10\log_{10}\left(\frac{\int_{cf(i-0.5)}^{cf(i+0.5)} |Q_\Psi(x,f)|^2 df}{cf(i+0.5) - cf(i-0.5)}\right), \quad (3)$$

where $cf(i)$ is the center frequency of $\mathrm{ERB_N}$ band i.
A first factor D_1 is defined as the standard deviation of excitation levels across $\mathrm{ERB_N}$ bands:

$$D_1(x) = \sigma(W(i) \times EQ_i(x)), \quad (4)$$

where $W(i)$ are weights that are applied to account for the limited importance of certain frequency bands in the coloration estimation (below $\sim 100\ Hz$ and above $\sim 10\ kHz$).
A second factor D_2 is defined as the standard deviation of the difference between excitation levels in successive $\mathrm{ERB_N}$ bands:

$$D_2(x) = \sigma(W(i) \times (EQ_{i+1}(x) - EQ_i(x))). \quad (5)$$

D_1 allows to observe general variations around the mean difference across frequency whereas D_2 estimates spectral ripple density. The overall weighted excitation pattern difference D is used to measure the coloration introduced by an electroacoustical system. It is defined as a weighted combination of D_1 and D_2:

$$D(x) = w \times D_1(x) + (1-w) \times D_2(x), \quad (6)$$

where w is set to 0.4 [12].

Figure 9 shows D values estimated from the loudspeaker array response at 3 m for the synthesis of both sources. The upper left graph shows mean values of D calculated over all frequency bands. The upper right graph shows standard deviation of D across microphone positions. It can thus be seen that D values vary little across microphone positions. There is no significant difference between DD and FD filtering. These type of filtering however give lower D values than MEQ filtering. The smallest D values are obtained for ideal loudspeakers. For "real" transducers, they are lower for MAP than for electrodynamic loudspeakers when using MEQ filters but the opposite occurs for the DD and FD filters. However, these differences are very small (~ 0.1 dB).
Similar tendencies are observed for mean values of D

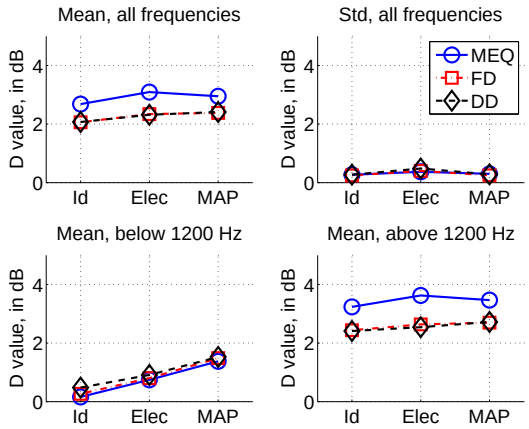


Fig. 9: D factor dependency on loudspeaker and filtering types, calculated from spatial responses obtained at 3 *m* from the loudspeaker array, Id: ideal omnidirectional loudspeakers, Elec: electrodynamic loudspeakers

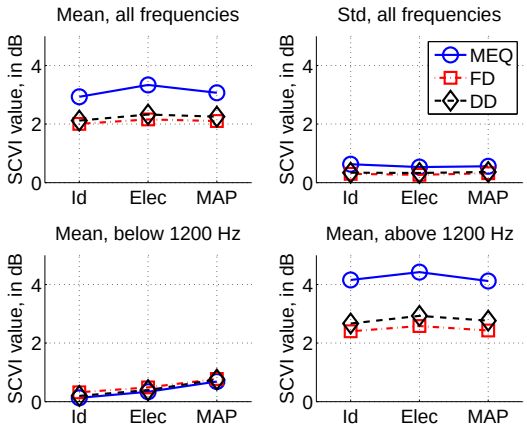


Fig. 10: SCVI dependency on loudspeaker and filtering types, calculated from spatial responses obtained at 3 *m* from the loudspeaker array, Id: ideal omnidirectional loudspeakers, Elec: electrodynamic loudspeakers

obtained by selecting frequency bands above 1200 Hz (lower right part of figure 9). The lower left part of figure 9 shows D values calculated using frequency bands below 1200 Hz. In this frequency range, ideal loudspeakers have very low D values. However, these values increase with electrodynamic ($D \sim 0.8$ dB) and MAP loudspeakers ($D \sim 1.4$ dB). This is due to their non ideal radiation characteristics that can only be partly compensated for using the multichannel equalization method [7].

3.3.2. Spatial color variation

Sound color variations may be experienced while moving the head or deambulating within the sound installation. This may happen in WFS installations, especially above the aliasing frequency where frequency/spatial patterns may appear (see figure 7(a)).
De Bruijn introduced the Spatial Color Variation Index (SCVI) [11]. We propose here a modified criterion expressed as:

$$SCVI(x) = \frac{1}{2 \times J} \sum_{j=-J}^{J} \sqrt{\sum_{i=1}^{I} (EQ_i(x) - EQ_i(x + j\Delta x))^2}. \quad (7)$$

We define $J = 2$ and $\Delta x = 10$ *cm*. This criterion is based on differences in $\mathrm{ERB_N}$ band i between excitation levels $EQ_i(x)$ at a position x and excitation levels $EQ_i(x + j\Delta x)$ at 4 positions on each side of the position x. The final criterion is then obtained by computing the mean value for all frequency bands.

Compared to de Bruijn's definition, this criterion extracts symmetrical positions on each side of the average listening position instead of relative positions in the same direction. It therefore avoids asymmetries of SCVI values along the x axis that could be noticed with the original criterion [11]. We also use the quality function instead of the original response of the loudspeaker array. This removes the influence of absolute level in the calculation that should not be accounted for in the calculation of a coloration index.

Figure 10 shows SCVI values estimated from the the loudspeaker array response at 3 *m* for both sources. figure the same as that of figure 9. We also observe similar tendencies:

- little or no position dependency,
- values are almost null at low frequencies and higher (SCVI ~ 3 dB) above 1200 Hz,
- diffusion reduces values of SCVI above 1200 Hz but there is no significant difference between DD and FD,

3.3.3. Discussion

The D and SCVI factors tend to show that there is a positive impact of diffusion on coloration above the aliasing

frequency. Diffusion may be introduced by natural properties of MAP loudspeakers or diffuse filtering. Only a certain amount of diffusion seems to be necessary. Using the FD filter which contains only "diffuse" energy does not decrease *D* or SCVI values compared to the DD filter.
However, the employed criteria are still questionable since they do not account for time and only partly for space (SVCI). Time related factors may be particularly important for WFS since the "effective length" of impulse responses may approach 10 to even 20 ms. Informal listening tests showed that using the FD filter on electrodynamic loudspeakers provides clearly noticeable coloration artifacts at high frequencies which tend to disappear almost completely when using DD filtering.
Similarly, above the aliasing frequency, the frequency response varies significantly given small changes of microphone position (5 to 10 *cm*)[13]. Audible and possibly disturbing coloration changes are experienced with MEQ filtering of the electrodynamic array while wandering in the sound installation if pink noise is used as an input signal. These coloration changes are reduced using the same type of filter on MAP loudspeakers and become almost inaudible using the DD filter on electrodynamic or MAP loudspeakers.

3.4. Spatial cues

The analysis proposed here considers a deviation estimation of time-related localisation cues. It is not intended as a localization model. Its role is rather to provide an evaluation of the deviation from "ideal" localization cues.
The evaluation is performed on the free field measurements using omnidirectional microphones. 94 pairs of microphones with 20 *cm* spacing are extracted from the microphone array. This corresponds to a simplified model of the localization cues provided to a listener facing the loudspeaker array.
Temporal localization cues are then estimated in $\mathrm{ERB_N}$ frequency bands from the loudspeaker array response on each microphone pair. The normalized interaural cross-correlation function at position x in frequency band i ($ICF_i(x)$) is calculated as:

$$ICF_i(x,\tau) = \frac{\int_0^{t_0} h_i(x+\Delta x,t)h_i(x-\Delta x,t+\tau)dt}{\sqrt{\int_0^{t_0}[h_i(x+\Delta x,t)]^2dt\int_0^{t_0}[h_i(x-\Delta x,t)]^2dt}} \quad (8)$$

where $\Delta x = 10\,cm$ and $h_i(x,t)$ is the band-pass filtered response in band i of the loudspeaker array response $h(x,t)$ using Gammatone filters [14]. In total 24 frequency channels are used to cover the audible range.
The Interaural Correlation coefficient $IC_i(x)$ can be extracted in frequency band i as:

$$IC_i(x) = \max_{\tau} ICF_i(x,\tau). \quad (9)$$

The corresponding Interaural Time Difference (ITD) in frequency band i is then defined as:

$$ITD_i(x) = \{\tau \,|\, ICF_i(x,\tau) = IC_i(x)\} \quad (10)$$

The latter is the only known human localization cue for low and mid frequencies (below 500 Hz) [15]. It is extracted from phase comparison between signals entering the left and the right ears. Above about 1200 Hz, $IC_i(x)$ and $ITD_i(x)$ are estimated from the signal envelope to account for fiber transduction in the auditory system. A classical way to extract the signal envelope and mimic the auditory system is to use half wave rectification and low-pass filtering (below 800 Hz) [15].

3.4.1. Interaural correlation

Figure 11 displays interaural correlation values below and above 1200 Hz (mean aliasing frequency). Mean values are computed from frequency bands below and above 1200 Hz for both virtual sources and all microphone pairs (upper left and right part of figure 11). Standard deviation values are computed across frequency and averaged for both virtual sources and all microphone pairs (lower left and right part of figure 11). Values are displayed for the 3 types of filtering and loudspeaker.
It can be seen that, as expected, the interaural correlation is very high below 1200 Hz ($IC \sim 1$) independently of the considered frequency band (std ~ 0). Above the aliasing frequency, *IC* values are still high. They are even close to 1 for the MEQ filter ($IC \sim 0.95$) and remain around 0.9 for the DD filter and 0.85 for the FD filter. Moreover, they vary little across frequency (std ~ 0.05).
It should be noted that mean IC values are similar for all loudspeaker types given a filtering type. Nonetheless it can be noted that for the MEQ filter, IC values increase from MAP (IC=0.94) to electrodynamic (IC = 0.95) to ideal (IC = 0.96).

3.4.2. Interaural time difference

Figure 12 displays ITD errors calculated below and above the corner frequency 1200 Hz. The ITD error is computed by subtracting ITD values estimated off of the ideal spatial response from ITD values estimated off of

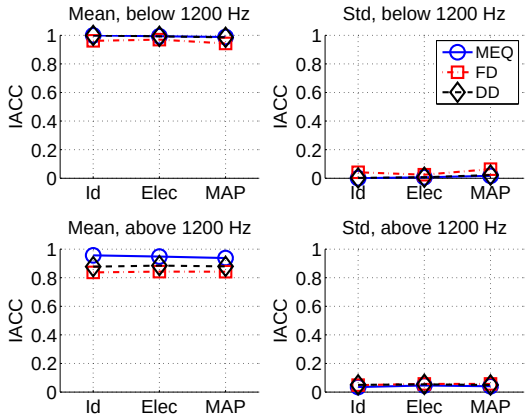


Fig. 11: IACC below and above 1200 Hz, Id: ideal omnidirectional loudspeakers, Elec: electrodynamic loudspeakers

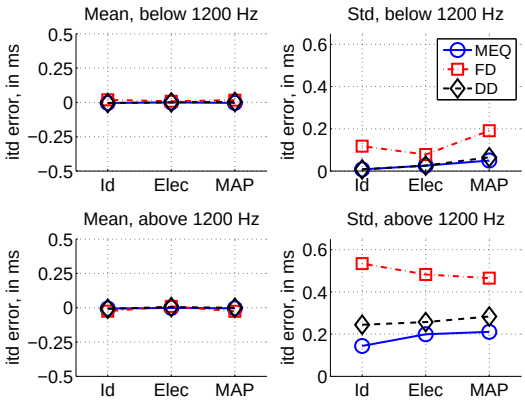


Fig. 12: ITD error below and above 1200 Hz, Id: ideal omnidirectional loudspeakers, Elec: electrodynamic loudspeakers

the response being considered. The definition of "mean" and "std" in figure 12 are the same as in the previous part. It can be seen that for both frequency regions (below and above 1200 Hz), the mean ITD error remains close to zero independently of filtering and loudspeaker type. The standard deviation value of ITD error at low frequencies is also close to 0. However, this value significantly differ from 0 above 1200 Hz. The standard deviation depend on loudspeaker and filtering type. Lower values of standard deviation are obtained for ideal omnidirectional loudspeakers (MEQ: 0.14 ms, DD: 0.24 ms) than for electrodynamic (MEQ: 0.2 ms, DD: 0.26 ms) and MAP loudspeakers (MEQ: 0.21 ms, DD: 0.28 ms).

3.4.3. Discussion

This analysis shows that, despite the complexity of the synthesized sound field above the aliasing frequency, the provided time-based localization cues remain at least partly consistent with target ones. IC values are high and mean ITD error is almost null. This can be attributed to the fact that the first peak of the impulse response which corresponds the loudspeaker located in the direction of the virtual source generates the first peak of the impulse response (verified at all listening positions if the virtual source is situated behind the loudspeaker array). For virtual sources within the listening environment (focused sources), the situation is different. In this case, the first loudspeaker contribution may come from the side of the loudspeaker array though at a fairly low level. More studies are required to evaluate these types of sources. The perceptual consequences of the variability of the estimated ITD at high frequencies have yet to be evaluated. They should be accompanied by an evaluation of Interaural Level Difference (ILD) in extended area using dummy head measurements. Early experiments were achieved in a reduced number of frequency bands in [6]. They show potential deviations of ILD for lateral listening positions. However, such inaccuracies may not be an impediment to localization of broadband signals exhibiting strong low frequency content. After Wightman and Kistler [16], low frequency ITD may dominate localization estimation for such signal. This is consistent with sound localization experiments conducted by Start. In his PhD [9], he showed a good match between localization performances for real and WFS reproduced sources using broadband white noise. In the same run of experiments, Start used high-pass filtered white noise (above 1200 Hz). He showed an increased mean error (5° instead of 2.5°) and standard deviation (3° instead of 1.5°). This error remains however limited and may be assimilated to an increased Auditory Source Width (ASW).
ASW is classically estimated using IC values. For a review, the reader is referred to [17]. In the present study, the observed deviation of IC from 1 within frequency bands is very limited considering the MEQ filter (IC $\sim$ 0.95) and appears *only* at high frequencies (above 1200 Hz). Such values are close to the Just Noticeable Difference (JND) and little is known about strictly high fre-

quency variations of IC. More variations can be noticed at high frequencies on the estimated ITD. These may also point to an enhanced ASW. Experiments on similar criteria were recently achieved by Hirvonen and Pulkki [18]. They divided a signal into frequency bands that were presented using horizontally distributed loudspeakers in an anechoic environment. These test signals elicited different ITD values in each frequency band with strong interaural correlation. However, their study is limited to low frequencies (below 1200 Hz) and would need to be extended to higher frequencies.

4. SUBJECTIVE COMPARISON

In this section, we present two subjective experiments that were carried out during August 2006 at IRCAM. The first experiment is a basic discrimination test on loudspeaker and filtering type. The second experiment is a pair comparison test on perceived distance which was quoted by the subjects as a first order attribute during the discrimination test.

4.1. Test setup

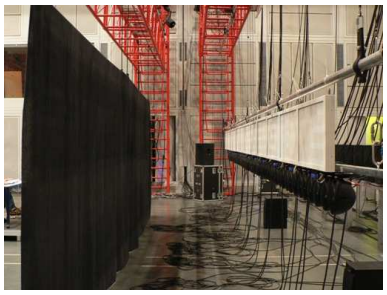


Fig. 13: Test setup

The test setup used for the listening test is composed of two arrays on top of each other (MAPs above, electrodynamic below, see figure 13). This arrangement is chosen so as to minimize potential discrimination of loudspeakers with height difference. They are positioned in the Espace de Projection, a 22(l)×15(w)×11(h) m^3 variable acoustic concert hall at IRCAM. All surfaces (periactes) are set to absorptive (walls and ceiling). The obtained reverberation time is therefore below 1 s at all frequencies. The choice of a 4.5 *m* distance between the subject and the loudspeakers is a tradeoff. On one hand, the distance should be sufficiently small distance so as to limit the influence of the room effect. On the other hand, it should be sufficiently large so as to avoid discrimination based on elevation difference. The apparent elevation difference of both loudspeaker arrays is 5 degrees which is below the localization blur for frontal position reported by Blauert in [15] ($\pm 9^\circ$).
Subjects are not centered but positioned 1.5 *m* to the left. An acoustically transparent curtain is placed between the subject and the loudspeaker array. To further limit visual feedback, all lights are dimmed. The main light source is the computer screen situated in front of the listener, approximately 40 *cm* below the ear level.

4.2. Discrimination test

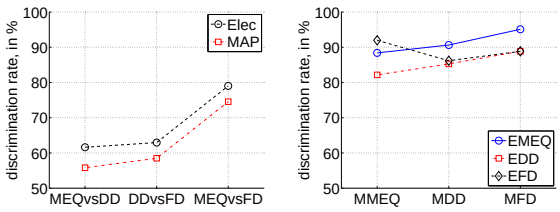
4.2.1. Test protocol

The subjective discrimination between loudspeakers and filtering types was evaluated using an ABX test. All 6 combinations of loudspeaker (E: electrodynamic, M:MAPs) and filtering types (MEQ, DD, FD) were used forming 15 comparison pairs. 60 triplets were then formed using 4 replicates according to the ABX test protocol (ABB, ABA, BAB, BAA).
Each triplet was presented at the two test virtual source positions depicted in figure 8 using two different sound materials (voice and guitar). The total of 240 triplets (60 "ABX" triplets × 2 positions × 2 sound materials) were presented in random order. The triplets were given only once, no repetition was possible. The presentation of one triplet took approximately 10 s (three times about 2.5 s, 1 s between configurations).
The 14 subjects completed the test in 1 hour on average. They had an initial training session with a maximum of 30 stimuli. Subjects were free to quit this training session before end.

4.2.2. Results

Figures 14(a) and 14(b) show mean discrimination rates for all stimuli pairs. Figure 14(a) shows within loudspeaker type results for different filtering. It can generally be seen that the discrimination rate is only slightly though significantly above chance level and that it is generally higher for electrodynamic than for MAP loudspeakers (about 5% higher). As expected, discrimination rate is higher for the pair "MEQ versus FD" than for the other two combinations.

Figure 14(b) shows discrimination rates for pairs having both different loudspeaker and filtering types. It can be seen that MAP and electrodynamic loudspeakers are clearly discriminated (80 to 95 % discrimination rate). Highest discrimination rates are obtained for MAP with FD filtering versus electrodynamic loudspeakers with



(a) Filtering dependency (MEQ versus DD, FD versus DD, MEQ versus DD)

(b) Filtering loudspeaker dependence

Fig. 14: ABX test discrimination rates (E: electrodynamic, M:MAP)

MEQ filtering. More generally, this configuration exhibits lowest discrimination rates against other filtering types on MAP loudspeakers.
A single way ANOVA did not show a significant effect of the sound material ($p \sim 0.2$) neither did it for the source position ($p \sim 0.6$).

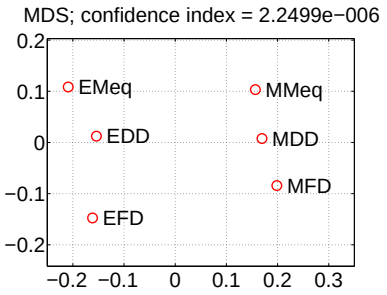


Fig. 15: MDS analysis of results for ABX test

A multidimensional scaling analysis was performed on the results using mdscale function of Matlab. Figure 15 displays the obtained space. The axes clearly define loudspeaker type (x axis) and filtering type (y axis). Stimuli are more spread along the loudspeaker axis than along the filtering axis, which is consistent with the results displayed in figures 14(a) and 14(b). The "EFD" stimulus (electrodynamic loudspeakers with FD filtering) appears slightly isolated since it always presents high discrimination rates.

4.2.3. Discussion

The general outcome of this test is that electrodynamic and MAP loudspeakers are well discriminated. A moderate amount of diffusion (DD filter) on electrodynamic loudspeakers reduces discrimination rate when compared to MAP loudspeakers which however remains above 80 %.
After the listening test, the subjects were asked to rank (from 1 to 5, most important to less important) an ensemble of five perceptual attributes to discriminate stimuli within the triplets. Average ranks given by subjects to these perceptual were computed. According to them, the most important attribute is distance (ranked 2.2 on average), then comes elevation (2.6) and coloration (2.9). The two last attributes (azimuth: 4 and ASW: 4.2) are ranked low. This may be due to the fact that the employed sound material are broadband. More experiments may be required on these specific attributes, possibly using more critical sound material (eg. high-pass filtered noise).
Even though the elevation difference is below 5°, the subjects were still discriminating the vertical positions of the loudspeaker arrays. This artificially increases the obtained discrimination rates between both loudspeaker types. Concerning coloration, the subjects may have used both coloration differences between loudspeaker types and coloration introduced by FD filter which may appear as severe.
The importance of distance confirmed impressions of the authors during early informal listening session. It was thus decided to conduct an additional listening test on this particular attribute.

4.3. Distance judgement

4.3.1. Test protocol

A pair comparison direct scaling method is used for the test on distance perception. Two successive stimuli are presented to the subjects. Their task is to indicate on a continuous scale if the second stimulus is closer, at the same distance, further than the first stimulus.
In this test, the FD filter was discarded since it was found to exhibit too much coloration that could be disturbing for subjects. This forms 4 stimuli and 6 pairs which are presented in both orders (12 and 21). The sound stimuli are rendered on both source positions.
A virtual room processor is used in order to elicit three levels of distance (close: no additional room effect, mid distance, far distance) [19]. The room effect is rendered using three virtual loudspeakers on the WFS array and 6 side and rear loudspeakers. The goal of this is to verify that the differences in distance perception between loudspeaker and/or filtering types remains consistent for various elicited distances. Only the guitar sound material was chosen since it was used in [19] to validate the virtual room model on a binaural setup.

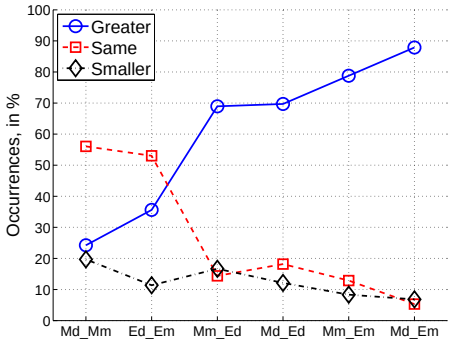


Fig. 16: Results for distance test. Occurrences of responses indicating greater, same, or smaller distance of stimulus 1 versus stimulus 2 (Em: electrodynamic MEQ filtering, Ed: electrodynamic DD filtering, Mm: MAP MEQ filtering, Md: MAP DD filtering).

A total of 72 pairs are presented only once in random order with no possible repetition. 11 subjects completed the test in an average of 20 minutes.

4.3.2. Results

The results of the test are simply analyzed by extracting the number of occurrences indicating a greater, smaller, or same distance for each of the 6 pairs. These are presented in figure 16. MAP loudspeakers are shown to significantly increase perceived distance compared to electrodynamic loudspeakers. Diffusion has a similar but less pronounced effect, especially considering the electrodynamic loudspeakers.
Single way analysis of variance did not show a significant influence of either pair order ($p > 0.9$) or elicited distance with virtual room effect ($p > 0.7$). Only a loose influence of source position can be noticed ($p \sim 0.2$).

4.3.3. Discussion

The perception of distance is classically linked to level, direct to reverberant energy ratio [19]. The more directive characteristics of electrodynamic loudspeakers at high frequencies may limit the interaction of the loudspeaker system with the listening room. A more complete analysis of produced room effect is thus necessary. In room measurements have been carried out but their analysis is beyond the scope of this article.
Another aspect concerns diffusion which seems to slightly increase the perceived distance. The objective data presented in section 3.4 show that diffusion reduces the precision of localization cues which may be interpreted as a factor increasing distance. However, this is noticeable only at high frequencies. More studies would be required on this aspect.

5. CONCLUSION

In this article, perceptual aspects of Wave Field Synthesis rendering using electrodynamic compared to MAP loudspeakers are investigated. It is hypothesized that the reduced spatial coherence that is observed with MAP loudspeakers may enhance the transparency of sound reproduction. Diffuse filtering for Wave Field Synthesis is then introduced so as to mimic the behavior of MAP on electrodynamic loudspeakers and potentially limit perceptual artifacts above the aliasing frequency. Objective criteria are used to evaluate perceptual dimensions linked to reproduction transparency (coloration and localisation cues). Diffusion is shown to potentially reduce coloration and sound color variation. Above the aliasing frequency, it is also shown that rather consistent localisation cues are available although the synthesized impulse response of the loudspeaker array is complex. Two subjective experiments showed that diffusion may account for some but not all differences between MAP and electrodynamic loudspeakers for WFS rendering. In particular, more studies are required to explain the increased perceived distance using MAP compared to electrodynamic loudspeakers.

6. REFERENCES

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

[2] M. Lombard and T. Ditton. At the heart of it all: The concept of presence. *Journal of Computer-Mediated Communication*, 3(2), September 1997.

[3] D. Begault. Auditory and non-auditory factors that potentially influence virtual acoustic imagery. In *Proc. AES 16th Int. Conf. on Spatial Sound Reproduction*, pages 13–26, Rovaniemi, Finland, 1999.

[4] 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, July/August 2004.

[5] V. Gontcharov and N. Hill. Diffusivity properties of distributed mode loudspeakers. In *108th Convention of the Audio Engineering Society*, February 2000.

[6] E. Corteel. *Caractrisation et Extensions de la Wave Field Synthesis en conditions relles d'coute*. PhD thesis, Universit de Paris VI, Paris, France, 2004. available at http://mediatheque.ircam.fr/articles/textes/Corteel04a/.

[7] E. Corteel. Equalization in extended area using multichannel inversion and wave field synthesis. *accepted for publication in Journal of the Audio Engineering Society*, 2006.

[8] B. J. C. Moore. *An Introduction to the Psychology of Hearing, 5th ed*. Academic Press, San Diego, CA, 2003.

[9] E. W. Start. *Direct Sound Enhancement by Wave Field Synthesis*. PhD thesis, TU Delft, Delft, Pays Bas, 1997.

[10] E. Corteel. On the use of irregularly spaced loudspeaker arrays for wave field synthesis, potential impact on spatial aliasing frequency. In *9th Int. Conference on Digital Audio Effects (DAFx-06)*, Montral, 2006.

[11] W. de Bruijn. *Application of Wave Field Synthesis in Videoconferencing*. PhD thesis, TU Delft, Delft, Pays Bas, 2004.

[12] B. J. C. Moore and C. T. Tan. Development and validation of a method for predicting the perceived naturalness of sounds subjected to spectral distortion. *Journal of the Audio Engineering Society*, 52(9):900–914, September 2004.

[13] E. Corteel, U. Horbach, and R. S. Pellegrini. Multichannel inverse filtering of multiexciter distributed mode loudspeaker for wave field synthesis. In *112th Convention of the Audio Engineering Society*, Munich, Allemagne, May 2002. Preprint Number 5611.

[14] M. Slaney. An efficient implementation of the patterson-holdsworth filter bank. Apple Computer Inc. - Technical Report number 35, 1993.

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

[16] F. L. Wightman and D. J. Kistler. The dominant role of low-frequency interaural time differences in sound localization. *The Journal of the Acoustical Society of America*, 91(3):1648–1641, March 1992.

[17] R. Mason, T. Brooks, and F. Rumsey. Frequency dependency of the relationship between perceived auditory source width and the interaural cross-correlation coefficient for time-invariant stimuli. *Journal of the Acoustical Society of America*, 117(3):1337–1350, March 2005.

[18] T. Hirvonen and V. Pulkki. Perception and analysis of selected auditory events with frequency-dependent directions. *Journal of the Audio Engineering Society*, 54(9):803–814, September 2006.

[19] R. Pellegrini and U. Horbach. Perception-based design of virtual rooms for sound reproduction. In *112th convention of the Audio Engineering Society*, Munich, Germany, May 2002.