Equalization in an Extended Area Using Multichannel Inversion and Wave Field Synthesis*

E. CORTEEL, *AES Associate Member*
(etienne.corteel@ircam.fr)

IRCAM, 75004 Paris, France
and
sonic emotion, CH-8154 Oberglatt, Switzerland

Wave field synthesis (WFS) targets the synthesis of the physical characteristics of a sound field in an extended listening area. This synthesis is, however, accompanied by noticeable reconstruction artifacts. They are due to both loudspeaker radiation characteristics and approximations to the underlying physical principles. These artifacts may introduce coloration, which must be compensated for over the entire listening area. Multichannel equalization techniques allow for the control of the sound field produced by a loudspeaker array at a limited number of positions. The control can be extended to a large portion of space by employing a new method that combines multichannel equalization with a linear microphone array–based description of the sound field and accounts for WFS rendering characteristics and limitations. The proposed method is evaluated using an objective coloration criterion. Its benefits compared to conventional equalization techniques are pointed out for both ideal omnidirectional loudspeakers and multi-actuator panels.

0 INTRODUCTION

Wave field synthesis (WFS) is a holophonic technique that relies on the reproduction of physical properties of sound fields in an extended listening area using linear arrays of loudspeakers [1], [2]. WFS generally considers the synthesis of "virtual" omnidirectional sources located in the horizontal plane. The corresponding loudspeaker driving signals are simply expressed as delayed and attenuated copies of a filtered version of the virtual source driving signal [3].

Approximations of the ideal underlying physical principles (Kirchhoff–Helmholtz and Rayleigh integrals) that are necessary in deriving a practical implementation introduce reproduction artifacts [3]–[6]. The use of loudspeakers that display nonideal directivity characteristics adds to these artifacts. Moreover, these artifacts depend on both the target virtual source characteristics and the listening position. They cause sound coloration and may introduce localization inaccuracies.

The goal of this paper is to present and validate, according to perceptually relevant criteria, an equalization technique adapted to WFS that reduces rendering artifacts in the entire listening area for any target virtual source. A typical WFS installation may consist of one or several horizontal arrays located in a horizontal plane comprising tens to hundreds of loudspeakers. In this paper the focus is on the compensation of free field rendering artifacts introduced by a linear array for the reproduction of omnidirectional virtual sources using WFS.

The loudspeaker array can be regarded as an acoustical aperture through which the incoming sound field (as emanating from a target sound source) propagates into an extended yet limited listening area. Simple geometrical considerations enable one to define a zone in which the virtual source is "visible" through the loudspeaker array. This "visibility zone" outlines the area in which the target sound field is properly reproduced [4], [5], [7]. Given the finite extension of the listening area and the loudspeaker array, sources can only be located in a limited zone so that they remain visible from within the entire listening area. This source visibility area is displayed in Fig. 1(a).

Considering the limitations of human sound localization capabilities, the source visibility area can be spatially sampled so as to define a finite ensemble of target virtual source positions [6]. The localization blur, or minimum audible angle, is defined as the just noticeable difference in azimuth perceptible by a listener. Blauert [8] reports a

*Manuscript received 2006 February 14; revised 2006 October 26 and November 7.

localization blur of ±3.6° for sources in front of the listener and of ±10° for lateral sources. However, for WFS applications it is not possible to distinguish between front or side since head orientation is arbitrary. The source grid displayed in Fig. 1(b) is created such that the minimum angle between any two sources taken within the source visibility area is inferior to ±2.5° as observed from any given listening position. This source grid defines a finite ensemble of elementary virtual sources to be synthesized with the prescribed WFS installation.

Conventional equalization techniques consider each loudspeaker of the multichannel reproduction system separately. They will be referred to as individual equalization techniques. They tend to compensate for the spatially averaged frequency response of each loudspeaker independently of the reproduction objective (target virtual source). This equalization does not account for specific radiation characteristics of each loudspeaker. Applied to WFS, these methods do not provide an accurate means for controlling the synthesized sound field over an extended listening area.

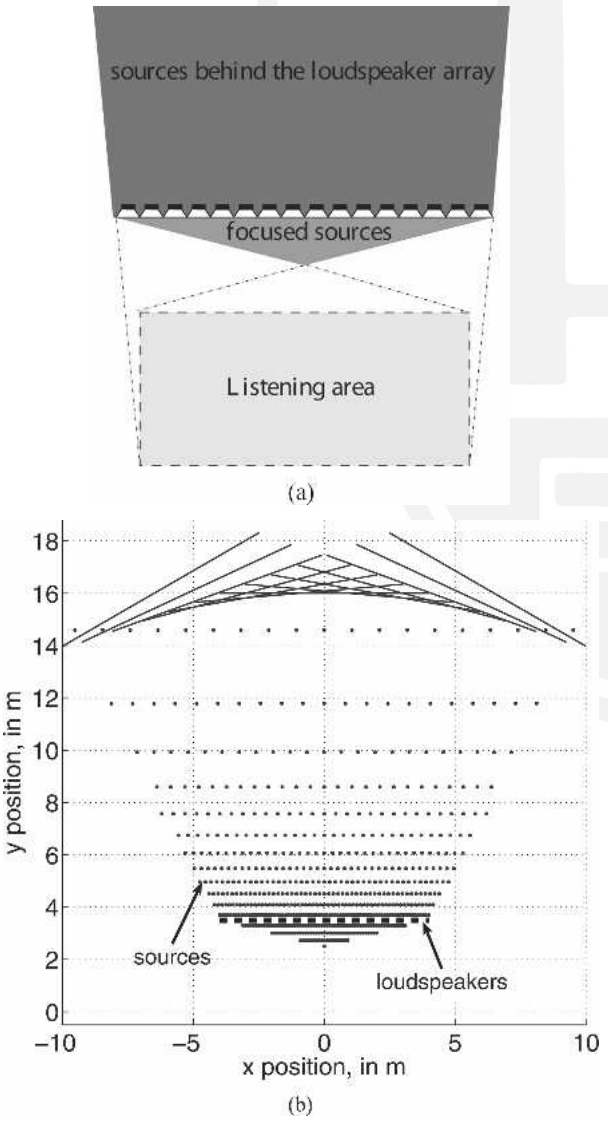

Fig. 1. Definition of source grid according to visibility criteria and localization blur. (a) Source visibility area. (b) Source grid.

Multichannel inverse filtering techniques [9] offer control of the sound field at a limited number of points in the listening space. Filters are calculated in order to minimize the reproduction error at the control points according to a *target.* The sound field radiated by the multichannel reproduction system is described by measuring loudspeaker impulse responses with microphones located at control points. These filtering techniques will be referred to as multichannel equalization. The solution produced by multichannel equalization techniques achieves both the synthesis of the reproduction objective (virtual source for WFS) and the compensation of reproduction artifacts in a unique process. Multichannel equalization techniques are blind processes, which are not constrained by the rules of any formal sound reproduction technique (stereo, Ambisonics, WFS, and so on). The only obvious connection with a given sound reproduction technique can be found in the loudspeaker arrangement and reproduction objectives. Multichannel equalization techniques ensure that the reproduction error reaches a minimum at the control points but cannot guarantee, as such, that the sound field is appropriate elsewhere. However, for WFS applications sound-field control must be effective over the entire listening area. It should not be restricted to a limited number of points.

Multichannel equalization has been applied to sound reproduction mostly in the case of Transaural[1] [10] or generalized Transaural [11] sound reproduction [9], [12]–[15]. This type of application requires that the listening area be restricted to the location of the listener's ears. However, in most applications listeners are likely to move away from the equalization points. Nelson et al. proposed to investigate the extent of the "equalization zone" [9] in which the synthesis error remains sufficiently low.

In [16] the author (together with Horbach and Pellegrini) proposed a method based partly on multichannel inversion (see also [17]). The method was applied to the reproduction objectives of WFS and aimed at controlling the free-field radiation patterns of multiexciter distributed-mode loudspeakers, now referred to as multi-actuator panels (MAPs) (see [18]). The method combined a sound-field description of the loudspeaker array obtained from a linear microphone array with a multichannel inversion in order to extend the size of the equalization zone. This method was later extended to the synthesis of directive virtual sources [19], [20] and to room compensation for WFS [21], [22]. A similar method for room compensation in WFS was presented by Spors et al. [23], [24]. It relies on circular microphone array measurements that are decomposed onto a plane wave or cylindrical harmonic basis [25]. Multichannel inversion is then performed in this transformed domain. The control remains efficient inside the circular array but may suffer from artifacts linked to both microphone and rendering system limitations [26].

In this paper a modified version of the technique presented in [16] is proposed. This new process combines

[1]Transaural is a registered trademark of Cooper Bauck Corporation

WFS in a simplified form and sound-field equalization. The resulting sound field displays a reduction of rendering artifacts in the entire listening area for all target virtual sources. This new technique can be regarded as an advanced calibration procedure in which filters that modify the loudspeaker driving signals are designed specifically for each source of the grid defined in Fig. 1(b).

Section 1 presents conditions for an efficient control within an extended area and proposes a modified multichannel equalization framework that can be applied to any given reproduction technique. Section 2 describes the artifacts introduced by WFS considering ideal loudspeakers. Possibilities and limitations inherent to WFS are also outlined in this section. In Section 3 the proposed multichannel equalization technique is described. Section 4 introduces compact objective criteria derived from perceptual studies by Moore and Tan [27] to evaluate sound coloration within the entire listening area using a reduced ensemble of representative virtual sources. Section 5 presents results of loudspeaker equalization considering a 48-channel test setup. The objective criteria of Section 4 are then used to evaluate two different loudspeaker technologies (ideal omnidirectional loudspeakers and multiactuator panels or MAPs [18]). The section also provides a study of free parameters of the multichannel equalization method. Section 6 discusses results and presents a practical implementation for real-time rendering.

0.1 List of Abbreviations

WFS	Wave field synthesis
MIMO	Multi-input multi-output (system)
ERB_N	(Normalized) equivalent rectangular bandwidth
Meq	Multichannel equalization

1 MULTICHANNEL EQUALIZATION FOR SOUND REPRODUCTION IN AN EXTENDED AREA

The application of multichannel inverse filtering for sound reproduction is made feasible by considering the sound reproduction system as a multiple-input multiple-output (MIMO) system. The MIMO system is described as loudspeaker driving functions (inputs) and signals captured by an ensemble of microphones (outputs) describing the radiated sound field in a limited number of locations (see Fig 2).

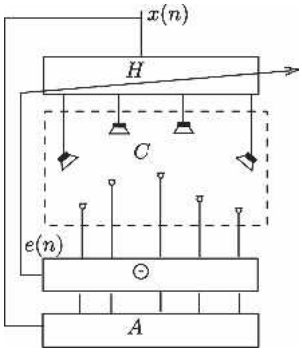

Fig. 2. Multichannel equalization applied to sound reproduction.

Filters $H(z)$ are calculated using multichannel inversion in order to minimize the reproduction error in reference to a given target. This requires the MIMO system to be identified by measuring the impulse responses of each loudspeaker on all microphones and organizing these into a loudspeaker–room transfer matrix $C(z)$.

The block diagram of the inversion process (see Fig. 3) is described in the following. A unique input signal $x(n)$ is filtered by the M-dimensional vector $H(z)$ and gives the M input signals $y_m(n)$ to the loudspeakers. The signals $y_m(n)$ are transmitted through the transfer function matrix $C(z)$, which outputs the L signals $z_l(n)$ which constitute the vector $z(n)$. The error associated with the channel l is the difference between the desired signal $d_l(n)$ and the output signal of the MIMO system $z_l(n)$. The desired signal $d_l(n)$ [forming $d(n)$] is specified by filtering the input signal $x(n)$ with the vector $A(z)$. The vector $A(z)$ corresponds to the acoustical transfer function associated with the desired sound field at the L control points.

Coefficients of the vector $H(z)$ are calculated in order to minimize the mean quadratic value of $e_l(n)$ in a multichannel inversion process. Three steps can therefore be identified:

1) Description of the MIMO system
2) Definition of the target sound field
3) Filter calculation

Applied to multichannel audio reproduction, multichannel inversion targets the control of the produced sound field at a limited number of points. However, one should notice that multichannel equalization only minimizes the synthesis error at a limited number of points. This does not guarantee that the sound field is correctly synthesized either at the control points or in the rest of the listening area.

The purpose of this section is to analyze how far multichannel equalization may provide an efficient control of a sound field produced by a loudspeaker system in an extended area.

1.1 Spatial Validity of the Solution

Supposing that the synthesized sound field fits the target at the control points perfectly, nothing ensures yet that it is correct in a larger portion of space. A necessary condition is that the loudspeaker radiation information collected on the specific microphone arrangement provide an unambiguous description in a larger subspace.

In the context of multichannel equalization, the sound field to be described is composed of the free-field radia-

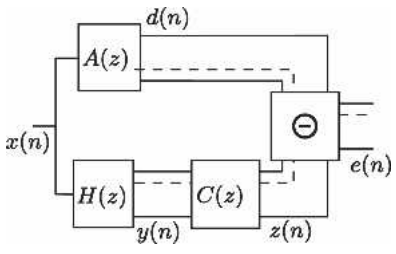

Fig. 3. Block diagram of inverse filtering process.

tion of the loudspeakers and/or the reflected contributions on the walls of the listening room. The description is realized in a listening area, outside the loudspeaker positions and any reflecting surface. Referring to acoustics principles, a sound field present in a subspace where there is no acoustical sources can be described either.

- As a pressure distribution in the complete subspace or
- As a pressure and/or pressure gradient distribution at the boundaries of that subspace (Kirchhoff–Helmholtz, Rayleigh integrals) or
- As a sum of independent solutions of the wave equation in a homogeneous medium (depending on the coordinate system chosen—spherical harmonics, cylindrical harmonics, plane waves).

A practical microphone setup allows one to capture the pressure or pressure gradient at a limited number of positions in space. Considering approaches 1) and 2), this restricts the extension of the possibly described subspace. It provides a spatially sampled description of the sound field. A corresponding Nyquist frequency can be derived depending on the microphone arrangement and the incoming sound field.

For approach 1) pressure microphones may be regularly spaced in a given volume. Note that using conventional Nyquist criteria for microphone arrays [28], a valid description up to 2 kHz for any incoming sound field in a volume of 1 cubic meter requires 1500 measurement positions.

Simplifications are obtained from approach 2). The Kirchhoff–Helmholtz integral states that the pressure sound field inside a closed surface ∂S is obtained from the knowledge of the pressure and the normal pressure gradient on ∂S. All acoustical sources must then be located outside of ∂S. For the same 1-cubic-meter portion of space, the number of required measurements is falling to 410. This approach has been proposed in a simplified form by Ise [29] and is known as boundary sound control (BSC). Ideally the microphone setup should cover a closed boundary (two dimensional surface) that limits a three-dimensional reproduction subspace. However, in [29] the proposed microphone setup is restricted to a rectangular distribution in the horizontal plane. This provides a description of the sound field that is valid in two dimensions only, assuming that the sound field is independent of height. However, this principle is also meant to compensate for listening room acoustics (room reflections, reverberation) which cannot be considered as a two-dimensional sound field. Therefore BSC, as described in [29], may not guarantee an exact reconstruction in the whole listening area.

In the limit case, where the closed surface is expanded to an infinite plane and the acoustical sources are located on the same side of that plane, the Rayleigh 1 integral applies. It states that the pressure distribution on the other side of the plane is uniquely derived from the knowledge of the pressure distribution on ∂S. An infinite number of measurements is a priori required, but a simplified approach will be proposed in Section 3.1.

Considering approach 3), any sound field is described from its decomposition into an infinite number of basic radiation elements. Practical solutions involve a specific microphone arrangement using conventional cells (such as omni, cardiod, figure of 8) and additional signal processing to decompose the captured sound field into a limited set of elementary solutions of the wave equation.

The area in which the sound field is properly described by such microphone arrays depends on many parameters, such as frequency, number of elementary solutions of the wave equation considered, microphone cell arrangement and directivity characteristics, and incoming sound field (see [30] for cylindrical harmonics). Available microphone array technologies are:

- Two- or three-dimensional, first-order (sound-field microphone [31]) or higher order Ambisonics microphones [32], [33]
- Linear or circular microphone arrays for plane wave decomposition [34], [25].

Multichannel inversion based on such a radiation description requires that both loudspeaker radiation and target be described on similar microphone arrangements and transferred into the same subset of elementary solutions of the wave equation. Spors et al. [24] proposed such an approach for the compensation of listening room compensation for WFS. They consider horizontal circular loudspeaker and microphone arrangements. It must be noted, however, that a circular microphone arrangement is an extreme case of a sampled cylindrical microphone array for cylindrical harmonics decomposition. Therefore all elevated reflected components are aliased on the two-dimensional description and may introduce artifacts during the inversion process [26].

1.2 Synthesis of Target at the Control Points

In the following the goal of multichannel inversion is to enable an independent control of the outputs of the MIMO system. From this point any desired signal can be synthesized at the output of the system. The invertibility of the loudspeaker–room transfer matrix is considered from a numerical approach. The calculation of the filters is then achieved using various filter design algorithms in the time or frequency domain. An alternative physical approach that is proposed is based on the capabilities of the loudspeaker rendering system to synthesize the target.

Numerical Approach From a numerical point of view any desired signals are obtained at the output of the MIMO system if the matrix C is invertible. In the general case, the optimum filters $\boldsymbol{H}_0$ may be calculated in the frequency domain as

$$\boldsymbol{H}_0(\omega) = [\boldsymbol{C}^{*\mathrm{T}}(\omega)\boldsymbol{C}(\omega)]^{-1}\boldsymbol{C}^{*\mathrm{T}}(\omega)\boldsymbol{A}(\omega) \quad (1)$$

where $\boldsymbol{C}^{*T}(\omega)[\boldsymbol{C}^{*T}(\omega)\boldsymbol{C}(\omega)]^{-1}$ is the pseudo inverse of $\boldsymbol{C}(\omega)$.

The multi-input multi-output inverse theorem (MINT), introduced by Miyoshi and Kaneda [12] in 1988, provides conditions for the invertibility of a loudspeaker–room transfer matrix. It states that an exact solution can be found if the number of loudspeakers involved exceeds the number of microphones and measured impulse responses do not have common zeros. The required filter length is high (several times the length of the measured impulse responses) and increases with the number of loudspeakers.

More generally, the "effort" necessary to invert a loudspeaker–room transfer function matrix is estimated by calculating the spreading of singular values of the matrix across frequency [35]. The conditioning number (ratio between largest and smallest singular values) provides such an estimate. When this number is infinite, the matrix is called ill-conditioned and no proper inverse can be found. For example, this is the case if all loudspeakers have common zeros.

Filter Design Algorithms Multichannel inversion can simply be achieved in the frequency domain using Eq. (1). In order to avoid ill-conditioning problems, Kirkeby et al. [14] introduced regularization techniques to add energy artificially at problematic frequencies. The inversion process becomes

$$\boldsymbol{H}_{0,\mathrm{reg}}(\omega) = [\boldsymbol{C}^{*\mathrm{T}}(\omega)\boldsymbol{C}(\omega) + \gamma \boldsymbol{B}^{*\mathrm{T}}(\omega)\boldsymbol{B}(\omega)]^{-1}\boldsymbol{C}^{*\mathrm{T}}(\omega)\boldsymbol{A}(\omega) \tag{2}$$

where $\boldsymbol{B}(\omega)$ is a frequency-dependent regularization matrix and γ a regularization gain. This technique prevents high Q resonances from appearing in the filters that may overload the loudspeakers at certain frequencies. The side effect of regularization is introducing errors in the inversion process. The regularization gain γ is tuned to avoid large errors while improving the invertibility of the system.

Filter calculation in the frequency domain appears to be a simple and cost-efficient method. The number of frequencies N_{inv} on which the inversion is performed corresponds to the number of points used for the discrete fourier transform (DFT). Considering that impulse responses in C have N_C points, the effective length of the corresponding filter N_H should be limited to $N_{\mathrm{inv}} - N_C + 1$. However, after applying the inverse DFT, significant contributions may appear in the filters at all N_{inv} points if the chosen filter length is not sufficient. This phenomenon is described by Norcross et al. [36] in the case of a unique loudspeaker equalization. It is due to the circular nature of the DFT and is known as wrapping or blocking effect. The filters obtained therefore contain artifacts that degrade the output signal quality. These can only be reduced by increasing the filter length N_H, thus increasing the number of frequencies N_{inv}. To limit the filter length, one should then window out a part of the calculated filters, potentially introducing errors.

Nelson et al. [9] described the error minimization process in the time domain. They derive a direct calculation of the filters based on the inversion of a matrix of unpractical size. As an alternative, they propose to use a multichannel adaptive filtering algorithm [least mean square (LMS) in their case] to derive iteratively filters that minimize the synthesis error. The main advantage of time-domain-based calculation is the possibility to derive filters that minimize the synthesis error for a given filter length without wrapping effects.

The proposed multichannel LMS algorithm is known to suffer from low convergence speed and may provide nonoptimum solutions (local instead of absolute minima of the error). Improvements in convergence speed and accuracy are obtained from multichannel versions of recursive least square (RLS) or affine projection algorithm (APA). Modified and fast versions of the APA have been introduced recently in the context of multichannel equalization for sound reproduction [16], [37]. They are known to present a good tradeoff between convergence speed and computational complexity compared to multichannel RLS algorithms.

Time-domain inversion using multichannel modified versions of the APA algorithm remains significatively more costly than frequency-based inversion, even if the number of points N_{inv} for the DFT is high. However, the filter design algorithm is usually performed in an off-line process. Computational constraints are thus lower than for real-time rendering, where the input signal is convolved with the calculated filters. In the latter case the filter length remains critical because it determines the required real-time processing power and possibly the input–output latency for frequency-domain-based convolution (such as the overlap–add method).

Physical Approach In the numerical approach, multichannel equalization can be compared to a black box. The filters are "simply" derived using multichannel inversion of the system to approach target output signals. Following MINT, an inversion is only possible if the number of control points is lower than the number of loudspeakers. However, the goal of multichannel equalization for the compensation of rendering artifacts is not the independent control of each output (control points) but the synthesis of a coherent sound field in an extended portion of space.

Targeting an effective equalization in an extended area, the control points should cover any position of the listening area. The numerical approach therefore fails to provide indications of the possibility to equalize the system in an extended area.

Nelson et al. considered the extension of the equalization zone, where the control of the sound field remains efficient for a simple case where only two loudspeakers and two microphones are used [38]. They report equalization zones around the control points on the order of a wavelength. However, they show that for some specific configurations of the target source, loudspeakers, and microphones, the equalization zone extends further away (microphones centered toward the loudspeakers, target source close to one of the loudspeakers). In these configurations the loudspeaker setup fits the needs of both target listening area and virtual source.

Sarris et al. [39], [40], following the work of Santillàn [41], proposed a multichannel equalization method to

compensate for listening room acoustics in an extended area using various loudspeaker setups controlled on a planar microphone array. They show in [40] that loudspeakers placed at the corners of the room are best suited to control the listening room acoustics. These are positions where the modes of the room have maximum amplitude and can be controlled best. In both example the efficiency of the multichannel equalization process is improved by considering a target that fits the loudspeaker arrangement and vice versa.

1.3 Application to a Sound-Reproduction Technique

Multichannel equalization in its basic formulation does not provide evidence of the possibility of controlling a sound field emitted by a loudspeaker ensemble in an extended area. However, this may occur if an appropriate microphone arrangement (sound-field description) is used and if the target is synthesized at the control points (see Section 1.1). A physical analysis of the loudspeaker arrangement and an appropriate target may provide indications about the feasibility of the control.

In the present paper, the aim of multichannel equalization is to account for loudspeaker deficiencies in the context of a specific rendering technique (WFS). Such a rendering technique defines a necessary loudspeaker arrangement and associated driving signals to synthesize a given target in a listening area, considering that loudspeakers have ideal radiation characteristics. The driving signals for each loudspeaker m correspond to gained, delayed, or more generally filtered [$k_m^{\psi}(z)$ in Fig. 4] version of an input signal with respect to the loudspeaker position and the target ψ.

A modified multichannel inversion scheme is proposed. It accounts for the sound-reproduction technique considered (see Fig 4). Modified filters $h_m^{\psi}(z)$ are calculated to minimize the error $e_l(z)$ for every output l,

$$e_l(z) = \sum_{m=1}^{M} \widetilde{h_m^{\Psi}}(z) k_m^{\Psi}(z) c_{lm}^{\Psi}(z). \tag{3}$$

Defining a modified matrix $\tilde{C}(z)$ such that

$$\widetilde{c_{ml}^{\Psi}}(z) = k_m^{\Psi}(z) c_{ml}(z) \tag{4}$$

the error to be minimized is given by

$$e_l(z) = \sum_{m=1}^{M} \widetilde{h_m^{\Psi}}(z)\, \widetilde{c_{ml}^{\Psi}}(z) \tag{5}$$

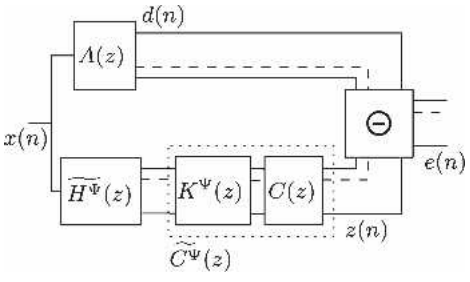

Fig. 4. Block diagram of modified inverse filtering process.

which rewrites as a conventional multichannel inversion problem.

The modified MIMO system to be taken into account is therefore described by $C^{\psi}(z)$. This step allows one to achieve a separation between the synthesis of the target sound field and the multichannel equalization. Therefore the latter focuses only on the compensation of reproduction artifacts (reproduction technique artifacts, loudspeaker frequency response, and directivities, and possibly listening room acoustics). The microphone arrangement should be defined according to the extent of the listening area covered by the sound-reproduction technique associated with the loudspeaker arrangement for each reproduction target. The target should then be defined by accounting for physical limitations of the rendering technique due to necessary practical simplifications. In the next section such an analysis is provided for WFS.

2 WAVE FIELD SYNTHESIS AND ITS LIMITATIONS

WFS is derived from the Huygens principle and its mathematical formulations and extensions (Kirchhoff–Helmholtz and Rayleigh integrals). This theoretical framework guarantees the exact synthesis of a target sound field in a reproduction subspace. Both integrals rely on continuous distributions of so-called secondary sources on a closed surface around the reproduction subspace (Kirchhoff–Helmholtz) or an infinite plane (Rayleigh 1 and 2).

The original formulation of WFS was proposed in the late 1980s by the Delft University of Technology [1]. It allows for the synthesis of virtual omnidirectional sources using a finite number of regularly spaced loudspeakers. Simplifications of the required loudspeaker array geometry are used to derive driving functions that depend mainly on source–loudspeaker relative positions. The latter introduces impairments in the reproduced wave field. Some may be compensated by filtering techniques. Some are physical limitations inherent in the loudspeaker geometry.

2.1 Derivation of Wave Field Synthesis

WFS is generally based on two assumptions: 1) Sources and listeners are located in the same horizontal plane, and 2) The target sound field emanates from a point source located at a given position with omnidirectional directivity characteristics.

The first assumption allows simplifying the required loudspeaker geometry. Using the second assumption, the pressure and the normal gradient radiated by the source can be estimated at any point in space.

The Rayleigh 1 integral [see Eq (6)] states that the synthesis of a sound field emitted by an ensemble of sources located in a half-space Ω_{ψ} can be synthesized in the other half-space Ω_R using a continuous distribution of ideal omnidirectional sources located on an infinite plane ($\delta\Omega$ separating Ω_S and Ω_R) driven by the normal pressure gradient emitted at their positions on $\delta\Omega$ (see Fig. 5). Con-

sidering that a unique source ψ creates the target sound field, the pressure caused by ψ at a position R in Ω_R is

$$p_\Psi(r_R) = -2 \int_{\partial\Omega} \frac{e^{-jk\Delta r}}{4\pi\Delta r} \nabla p_\Psi(\boldsymbol{r}_S) \cdot \boldsymbol{n}\, dS. \quad (6)$$

If Ψ is an ideal omnidirectional source, the normal pressure gradient $\nabla p_\Psi(\boldsymbol{r}_S) \cdot \boldsymbol{n}$ at position r_S on $\delta\Omega$ is expressed as

$$\nabla p_\Psi(\boldsymbol{r}_S) \cdot \boldsymbol{n} = S(\omega) \frac{\boldsymbol{r}}{r} \cdot \boldsymbol{n} \frac{1 + jkr}{r} \frac{e^{-jkr}}{4\pi r} \quad (7)$$

where $\boldsymbol{r} = \boldsymbol{r}_\Psi - \boldsymbol{r}_S$, r is the norm of $\boldsymbol{r}$, and $S(\omega)$ is the driving signal of the source Ψ. Assuming that Ψ is located in the far field of any point on the surface $\partial\Omega$ ($kr \gg 1$), Eq. (7) can be simplified,

$$\nabla p_\Psi(\boldsymbol{r}_S) \cdot \boldsymbol{n} = S(\omega) \cos\theta_\Psi jk \frac{e^{-jkr}}{4\pi r} \quad (8)$$

where

$$\cos\theta_\Psi = \frac{\boldsymbol{r}}{r} \cdot \boldsymbol{n}. \quad (9)$$

The WFS driving functions are then derived considering the radiation of each vertical line C of $\partial\Omega$. From the fact that both ψ and R are located in the same horizontal plane, the so-called stationary phase approximation shows that the main contribution of C to the synthesized sound field observed from R originates from the secondary source at the intersection of C and the horizontal plane. The planar secondary source distribution can thus be simplified as a linear distribution in the horizontal plane.

The WFS driving function $U_\psi(x_L, k)$ [3] of an ideal omnidirectional loudspeaker located at (x_L, y_L, z_L) to synthesize a virtual source ψ using the stationary phase approximation is expressed as

$$U(x_L, k) = S(\omega) \sqrt{\frac{k}{2\pi}}\, g_\Psi(y_{R_{\text{ref}}}) \cos\theta_0 \frac{e^{-j(kr_0 - \pi/4)}}{\sqrt{r_0}} \quad (10)$$

where r_0 and θ_0 are the values taken by r and θ_Ψ for the position of the secondary source (x_L, y_L, z_L) considered,

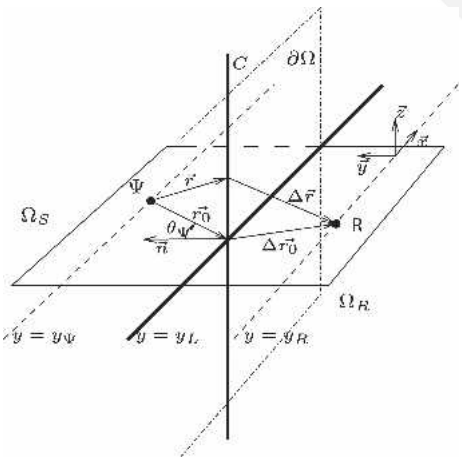

Fig. 5. Derivation of WFS driving signals from Rayleigh 1 integral.

g_Ψ is a factor introduced by the stationary phase approximation that allows to compensate for the level inaccuracies due to the simplification of the geometry of the loudspeaker array, such as

$$g_\Psi(y_R) = \sqrt{\frac{|y_R - y_L|}{|y_R - y_\Psi|}}. \quad (11)$$

However, the level is correct only at a reference listening depth $y_{R_{\text{ref}}}$.

The WFS driving function of a loudspeaker located at x_L can be regarded as a modified version of the source driving signal using

- Delay corresponding to the propagation time between ψ and r_L
- Gain corresponding to the natural attenuation of the source and compensation of the missing part of the planar secondary source distribution
- A 3-dB per octave filter ($\sqrt{k}$) accounting for the far-field estimation of the pressure gradient (+6 dB per octave) and the missing part of the planar secondary source distribution (−3 dB per octave)
- A $-\pi/4$ phase shift that can be described as a frequency-dependent delay of $-T/8$, where T is the corresponding period, this means that low frequencies have to be emitted before high frequencies to compensate, once again, for the missing part of the secondary source distribution.

Note that the two last elements do not depend on the loudspeaker position. They can be seen as an equalization filter that aims at compensating the missing parts of the planar array. Mostly delays are responsible for the wavefront curvature. They "shape" the emitted wavefront by the loudspeaker array while gains maintain an as correct as possible sound-field attenuation. Similar driving functions can be derived for focused sources (sources located in Ω_R) by considering the time reversibility of the acoustic equations [7].

2.2 Limitations in Practical Configurations

In the derivation of WFS driving functions, the required loudspeaker geometry is a continuous infinite line. For practical implementations this geometry is restricted to a finite number of closely spaced loudspeakers on a limited segment. These additional simplifications introduce limitations in the reproduced sound field that have to be accounted for while attempting to control the sound field using multichannel equalization.

Sound Field Attenuation The simplified secondary source geometry modifies the attenuation of the synthesized sound field. The $g\psi$ factor that appears in the generalized WFS driving functions allows one to maintain an accurate level on a line parallel to the loudspeaker array. Such a line is identified as $y = y_{R_{\text{ref}}}$, $z = 0$ (horizontal plane), and referred to as average listening depth,

$$g_\Psi(y_R) = \sqrt{\frac{\Delta r_0}{r_0 + \Delta r_0}} = \sqrt{\frac{|y_R - y_L|}{|y_R - y_\Psi|}}. \quad (12)$$

Outside of this line the level of the sound field at position r_R can be estimated using the stationary phase approximation along the x dimension [6]. The corresponding attenuation law Att_Ψ is expressed as

$$\mathrm{Att}_\Psi(r_R) = \sqrt{\frac{y_{R_{\mathrm{ref}}}}{y_R}} \sqrt{\frac{y_R + y_\Psi}{y_{R_{\mathrm{ref}}} + y_\Psi}} \frac{1}{4\pi d_\Psi^R}. \tag{13}$$

d_Ψ^R denotes the distance between the primary source Ψ_m and the listening position r_R. It appears as a combination of the natural attenuation of the target virtual source ($1/4\pi d_\Psi^R$) and the line array ($\sqrt{1/y_R}$) which can be physically compensated for only at the average listening depth. In the following, these attenuation characteristics are referred to as "WFS attenuation law."

Windowing, Diffraction For obvious practical reasons the length of the loudspeaker array must be restricted to fit in the listening room. The loudspeaker segment can be regarded as an acoustical window that limits the visibility of the virtual sound source to a restricted area.

As for any wave phenomenon propagating through an aperture, diffraction occurs. Diffraction waves are emitted both inside and outside the visibility area, resulting in oscillations in the frequency response. This effect can be reduced simply by attenuating the driving signals of the loudspeakers located at the extremities of the loudspeaker array. Start [4] proposes to use cosine–sine attenuation factors for the loudspeakers located at each side of the array for 10% of the total array length.

Fig. 6 displays level compensated frequency responses at different distances from a 6-m-long loudspeaker array synthesizing a centered virtual source located 6 m behind the loudspeaker array. The characteristics of the oscillations at mid and high frequencies depend on the listening position. These are reduced by attenuating side loudspeakers, but only above 300 or 500 Hz, depending on the microphone position [see Fig. 6(c) and (d)]. Below 100 Hz a rolloff can be observed. The associated corner frequency tends to increase with the distance from the loudspeaker array [6]. This can be compared to the classical limit distance between close-field and far-field attenuation from

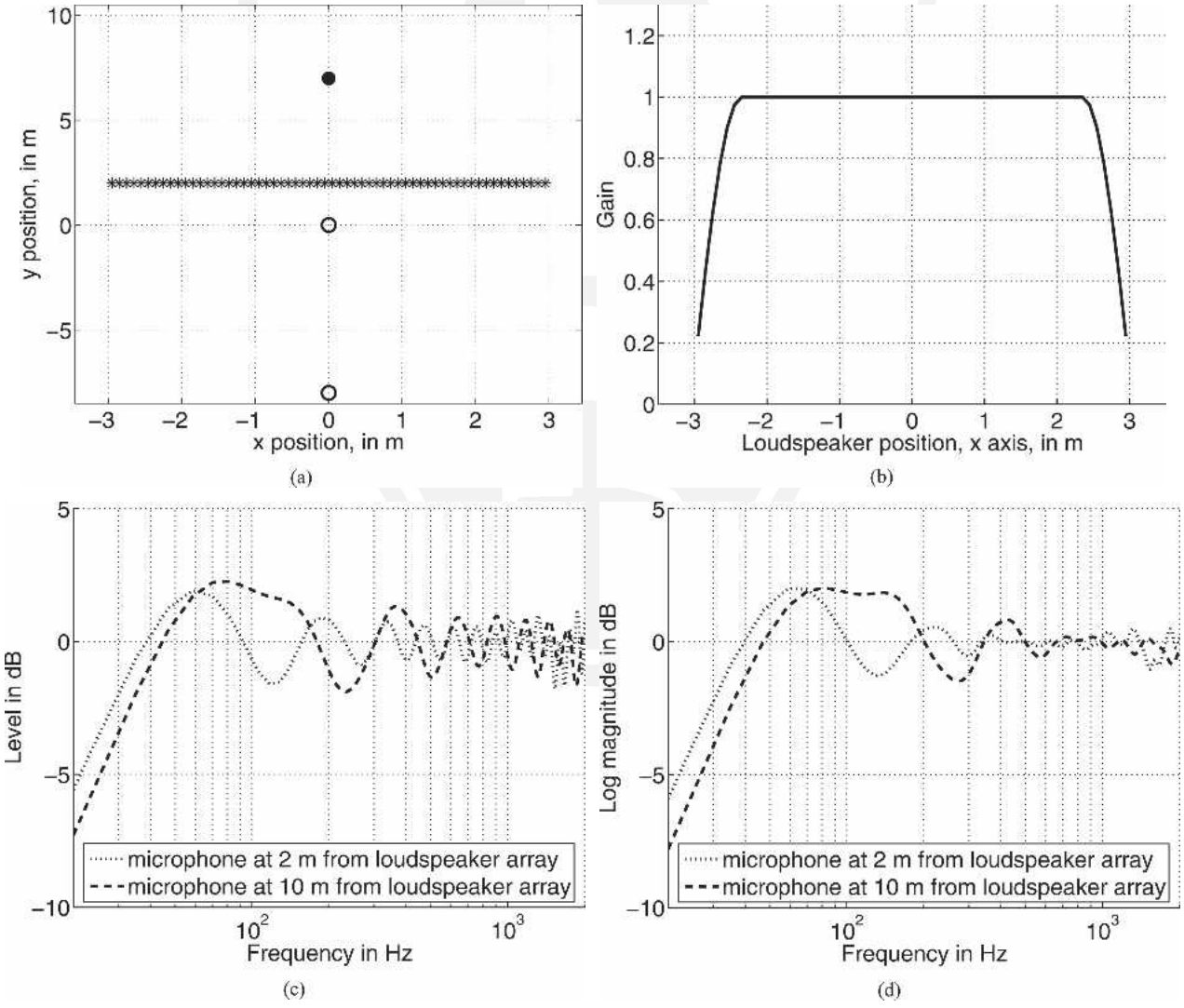

Fig. 6. Windowing effect on frequency response. (a) Top view of loudspeaker (*), microphone (○), and source (●) configuration. (b) Attenuation applied to loudspeakers for diffraction compensation. (c) Level-compensated frequency responses, dependency on listening position, without attenuation of side loudspeakers. (d) Level-compensated frequency responses, dependency on listening position, with attenuation of side loudspeakers.

the line source literature (see [42] for a review). This corresponds to the frequency-dependent limit distance for which contributions coming from all loudspeakers of the array are in phase. The sound field thus attenuates following a 6-dB per doubling distance law rather than the WFS attenuation law from Eq (13). The latter is observed mainly at low frequencies (< 100 Hz), which are, however, below the cutoff frequency of currently used loudspeakers for WFS.

Sampled Array Available WFS arrays use a limited number of regularly spaced loudspeakers. The corresponding spatial sampling process limits the accurate synthesis of the target sound field to a Nyquist frequency usually referred to as aliasing frequency. The aliasing frequency is known to depend only on the source position and loudspeaker spacing [4], [5].

Fig. 7 displays the full bandwidth response of the system of Fig. 6(a) (60 loudspeakers, 100-mm spacing, virtual source centered toward the array, 5 m behind). In the frequency domain [see Fig. 7(a)] the aliasing creates a very irregular response starting from the aliasing frequency. It can be seen that the aliasing frequency increases with the microphone distance from the loudspeaker array (about 2 kHz at 2 m and 3 kHz at 10 m).

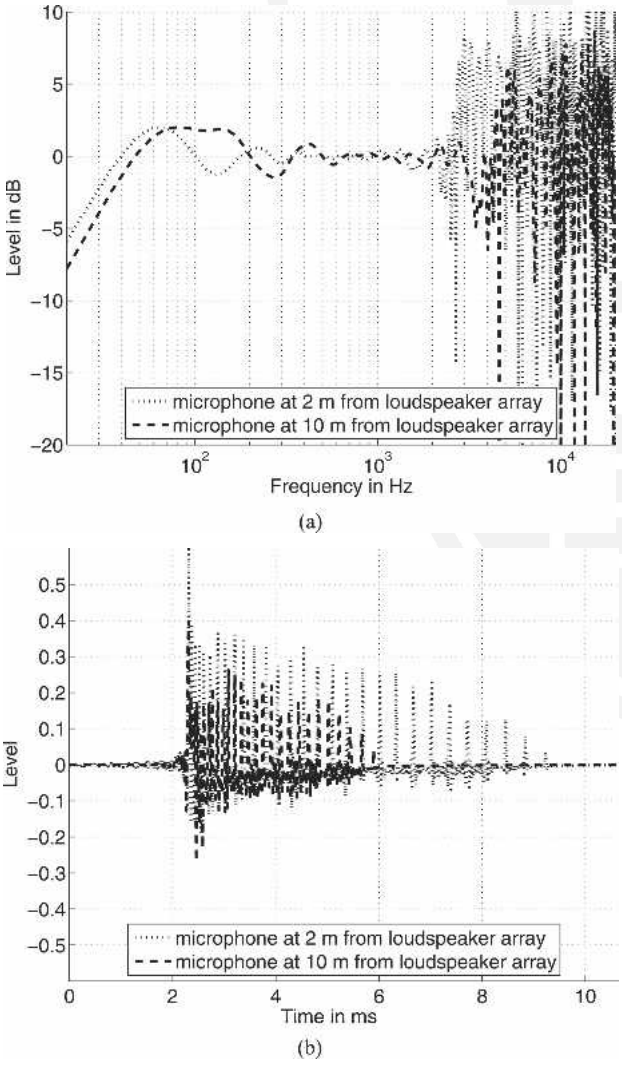

Fig. 7. Aliasing effect, configuration of Fig. 6(a). (a) Level-compensated frequency responses, dependency on listening position. (b) Level/time-compensated impulse responses, dependency on listening position.

Impulse responses of the loudspeaker array measured at the close and far positions of the microphones are displayed in Fig. 7(b). The impulse responses were time aligned by removing the propagation time difference between both microphone positions. This enables one to compare directly the temporal distribution of both impulse responses. They show a clear first peak at the expected natural propagation time between the virtual source and the listening position. A series of peaks then corresponds to the individual contributions of the loudspeakers of the array. These individual contributions no longer fuse into a single peak above the aliasing frequency. The length of the nonnull portion of the impulse response appears to be shorter for the microphone at 10 m than for the microphone at 2 m [see Fig. 7(b)]. At 10-m distance the loudspeaker array is indeed seen smaller than at 2 m. Therefore individual loudspeaker contributions are more time coherent at 10-m distance than at 2-m distance.

An alternative formulation of the aliasing frequency can be given from an analysis of the wavefront forming in the time domain. It accounts for the listening position dependency that is due to the finite length of the loudspeaker array [6], [43]. The aliasing frequency at listening position $\boldsymbol{r}$ and for source ψ is expressed as

$$f_{\mathrm{al}}(\boldsymbol{r}, \Psi) = \frac{1}{\max_{l=1,\ldots,L-1} t_{l+1}(\boldsymbol{r}, \Psi) - t_l(\boldsymbol{r}, \Psi)} \tag{14}$$

where, $\max_{l=1,\ldots,L-1} t_{l+1}(\boldsymbol{r}, \Psi) - t_l(\boldsymbol{r}, \Psi)$ is the maximum time difference of the contributions coming from successive loudspeakers of the array at listening position $\boldsymbol{r}$.

Spatial aliasing frequency definitions in the literature use the spatial Fourier transform to describe the target sound field and the sampling process [5], [4]. They consider an infinitely long linear loudspeaker array. The aliasing frequency obtained does not depend on the listening position, which is not correct for finite-length loudspeaker arrays. While considering infinite-length loudspeaker arrays, Eq. (18) can be shown to provide the same results than the usual formula based on the spatial Fourier transform [6].

3 MULTICHANNEL EQUALIZATION FOR WAVE FIELD SYNTHESIS

In this section we present a method that uses multichannel equalization for WFS rendering in order to account for loudspeakers and rendering technique deficiencies. The three steps of multichannel inversion (MIMO system identification, target definition, filter calculation) are presented. Particularly highlighted are the free parameters of the method. The multichannel inversion is only performed below the aliasing frequency. A modified equalization is applied at high frequencies.

3.1 MIMO System Identification

The identification of the MIMO system is achieved by measuring the free-field impulse responses of each loud-

speaker on an ensemble of microphones. As stated in Section 1.1, the local measurements should correspond to an objective characterization of the radiation of each loudspeaker in the entire target listening area. We propose to rely on linear arrays using a limited number of regularly spaced microphones. The simplifications on the microphone array geometry can be compared to the simplifications used to derive WFS from similar physical principles (Rayleigh 2 integral).

1) The sound field has to be described mainly in the horizontal plane in which loudspeakers are located. Similarly to WFS, the main contribution to the radiation of the loudspeakers in the horizontal plane is the sound pressure measured at the intersection of the infinite plane ∂S and the horizontal plane.

2) Synthesizing a virtual source with a finite-length loudspeaker array introduces a windowing effect and limits the size of the listening area. Outside of the enlightened zone, only diffraction effects appear. The microphone array might thus be located entirely in this restricted area.

3) The limited number of loudspeakers introduces an aliasing effect above a certain frequency. The synthesized sound field exhibits aliased spatial contributions and cannot be controlled efficiently [6]. Therefore the radiation description is required only for low and mid frequencies, and the number of required microphones can be reduced.

All these simplifications do not provide a complete description of the radiated sound field in the listening area. However, it is a compromise that allows one to reduce the required amount of measurement while getting the main contributions to the sound field in the listening area.

The spatial sampling of the microphone array introduces some aliasing in the description. The associated Nyquist frequency is defined in the microphone array literature [28] as

$$f_{\text{al}}^{\text{Mes}} = \frac{c}{\Delta x(1 + \sin \theta_{\Psi,\max})} \tag{15}$$

where $\theta_{\Psi,\max}$ is the maximum incoming angle of any plane wave component of the captured sound field (see Fig. 8) and c is the speed of sound. This frequency should remain sufficiently high to exceed the maximum WFS aliasing frequency for any considered source and listening position.

3.2 Target Response

The target responses are defined according to the sound field emanating from a target virtual source captured by the microphone array. From the knowledge of source position, microphone position, and directivity characteristics, the target response can be simulated. In order to account for specific attenuation of the sound field radiated by a linear loudspeaker array, the target sound field is defined as the ideal response of the loudspeaker array for the synthesis of source Ψ using WFS. The target sound field is further defined by choosing a normalization point at which all sources have the same propagation delay and the same level. In the following, this point corresponds to the origin of the Cartesian system (see Fig. 9). The target impulse response at an omnidirectional microphone located at (x_M, y_M) considering a loudspeaker array L and a virtual source Ψ located at (x_ψ, y_ψ) is

$$A(x_M, y_M, t) = \sqrt{\frac{d_O^L}{d_M^L}} \sqrt{\frac{d_M^L + d_\Psi^L}{d_O^L + d_\Psi^L}} \frac{d_\Psi^O}{d_\Psi^M} \times \delta\left(t - \frac{d_\Psi^M - d_\Psi^O}{c} - \tau_{\text{eq}}\right) \tag{16}$$

where τ_{eq} is an additional delay in order to ensure that calculated filters are causal. In the following τ_{eq} is referred to as equalization delay.

3.3 Multichannel Inversion

Once the MIMO system is identified and the target defined, filters may be calculated from multichannel inversion using, for example, the multichannel MFAP algorithm [16]. However, a few additional steps are performed in order to adapt the multichannel inversion to WFS and ensure an accurate control outside the equalization points.

Loudspeaker and Microphone Selection A finite-length loudspeaker array enables the synthesis of the sound field associated with a virtual sound source in a limited portion of space (see Section 2.2). This is a physical limitation of the technique. Therefore it was chosen not to attempt any control outside the visibility area. For example, in the case of sources located to the side of the array, some microphones may be located outside of the visibility area (see Fig 10). They should be ignored in the multichannel inversion process.

In some situations the microphone array does not span the entire visibility area. This is the case for sources close to the loudspeaker array. A visibility area may be defined through the microphone array (see Fig. 10). It determines the portion of the loudspeaker array that directly contributes to the sound field in the area considered. An exclusive selection of these loudspeakers may speed up the calcula-

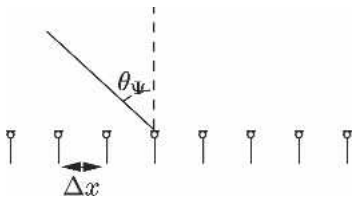

Fig. 8. Measurement aliasing description.

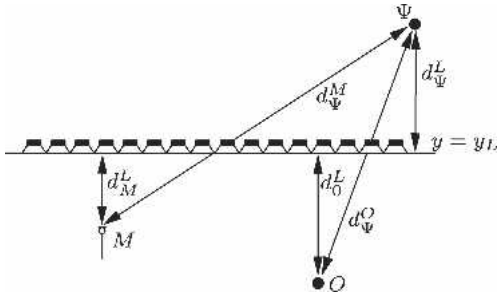

Fig. 9. Target sound-field definition and normalization.

tion of the equalization filters significantly during the multichannel inversion process. However, it would reduce the amount of available control parameters for equalization of the sound field. A tolerance distance $d_{\text{vis}}^{\text{ls}}$ is introduced to extend the selection of loudspeakers outside the direct visibility area (see Fig. 10). This appears as a second free parameter of the method.

Adaptation to WFS As proposed in section 1.3, WFS driving functions are applied to the matrix of impulse responses in order to separate wavefront forming and compensation of reproduction artifacts (equalization). A modified matrix of the impulse responses $\widetilde{C^{\Psi}}(t)$ is defined,

$$\widetilde{c_{ml}^{\Psi}}(t) = \delta(t - \tau_m^{\Psi}) * c_{ml}(t) \tag{17}$$

where $c_{ml}(t)$ is the impulse response of loudspeaker m measured by microphone l. τ_m^{Ψ} is the delay associated with loudspeaker m that appears in the WFS driving functions for the synthesis of the virtual source Ψ [see Eq. (10)]. * denotes the convolution operator.

The gain factor of Eq. (10) is not included since it tends to degrade $\widetilde{C^{\Psi}}$ conditioning, especially for sources close to the loudspeaker array. This determines the necessary effort to invert $\widetilde{C^{\Psi}}$ (see Section 1.2). This is illustrated in Fig. 11 in the case of a 48-channel ideal loudspeaker array for the reproduction of a source very close to the loudspeaker array. The original matrix C is obtained by simulating the impulse responses obtained on a 96-channel microphone array with 100-mm spacing at 2 m from the loudspeaker array. The modified $\widetilde{C^{\Psi}}$ matrix is obtained by

- Only delaying impulse responses of C with τ_m^{Ψ}
- Delaying impulse responses of C with τ_m^{Ψ} and applying WFS gains too.

Fig. 11(b) displays the conditioning number calculated between 100 and 2000 Hz for the three situations. It can be seen that the conditioning of the matrix is generally poor and even degrades when WFS gains are applied for the calculation of $\widetilde{C^{\Psi}}$.

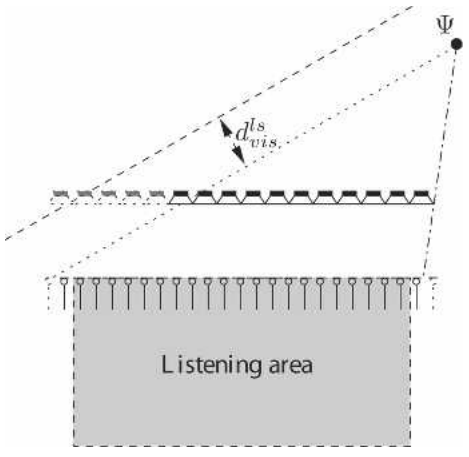

Fig. 10. Microphone and loudspeaker selection.

WFS driving functions consist of delayed and attenuated functions of a filtered ($\sqrt{jk}$) version of the input signal [see Eq. (10)]. Considering a 10-m-long loudspeaker array, maximum WFS delay differences between loudspeakers may approach 1000 taps at a 48-kHz sampling rate. Therefore a large number of taps of the filters calculated by multichannel inversion without preprocessing C would only account for the natural propagation of wavefronts and not for the compensation of rendering artifacts. The filter lengths L_{filt} obtained from multichannel inversion can be optimized by preprocessing measured impulse responses with WFS delays. All filters are thus time aligned and their main peak appears at τ_{eq} [see Eq. (16)].

The mean square criterion used for error minimization in the multichannel inversion process is time invariant. It spreads the remaining errors uniformly in the time domain before and after the main peak. However, listeners may be more sensitive to errors occurring in the time domain be-

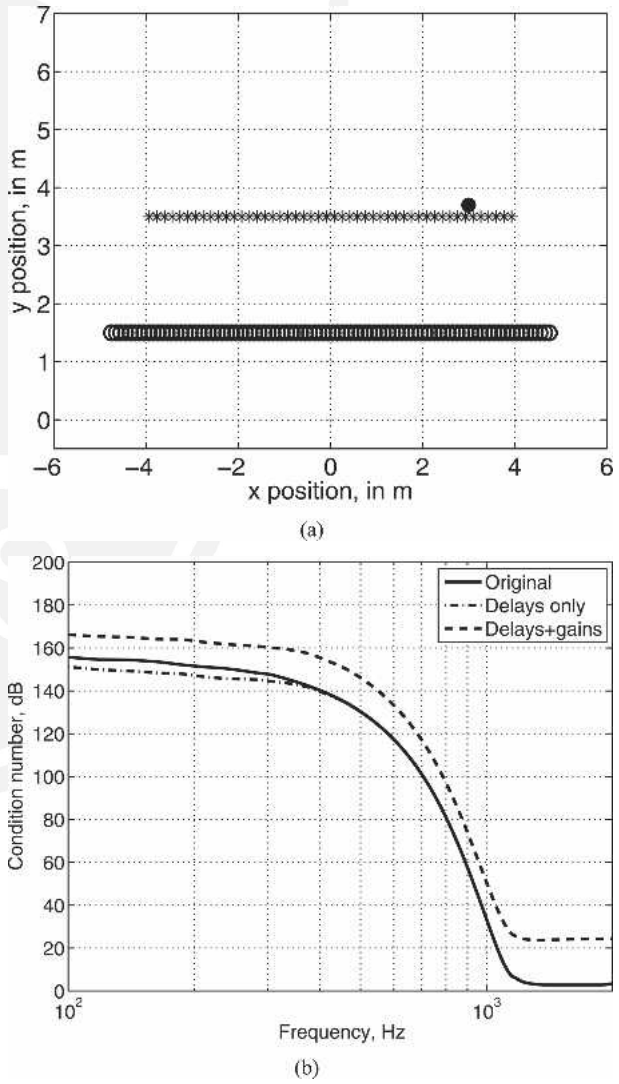

Fig. 11. C^{Ψ} matrix conditioning depending on type of preprocessing applied. (a) Top view of test configuration of loudspeakers (*), microphones (○), and virtual source (●). (b) Conditioning number calculated for C^{Ψ}. —— original (no preprocessing, C); –·– original with WFS delays applied; ––– original with WFS delays and gains.

fore the main peak rather than after, because of limited backward masking properties of human hearing (compared to forward masking; see [44]). All contributions of the filters before the main peak (preechoes) can be useful to equalize the responses of the loudspeakers but may also appear because of measurement errors, for example. After multichannel inversion, errors may thus be minimized, but with unnecessary contributions of the filters before the main peak from both a physical (measurement errors) and a perceptual (backward masking limitations) point of view. Therefore the delay τ_{eq} should be minimized to avoid equalization artifacts while maintaining an efficient equalization of the system.

Upper Calculation Frequency Spatial aliasing is a fundamental limitation of WFS. It is due to the spatial sampling of the originally required continuous loudspeaker distribution. The corresponding Nyquist frequency depends on source position, loudspeaker spacing, and, for a finite-length array, listening position [see Eq. (18)]. Above this Nyquist frequency the synthesized sound field is aliased and partly undetermined. Any attempt to achieve multichannel equalization may realize a local solution. The control can be shown to be valid at the control points only [6].

The formulation of the aliasing frequency for WFS is obtained by considering the spatial sampling as a temporal sampling of the wavefront synthesized by the loudspeaker array [see Eq. (18)]. Similarly it is possible to define a loudspeaker-related aliasing frequency $f_{\mathrm{al}}^{\mathrm{ls}}(\mathrm{ls}_l, \tau, \Psi)$ at position $\boldsymbol{r}$ using a similar temporal sampling criterion,

$$f_{\mathrm{al}}^{\mathrm{ls}}(\mathrm{ls}_l, \boldsymbol{r}, \Psi) = \frac{1}{\max\left|t_{l+1}(\boldsymbol{r}, \Psi) - t_l(\boldsymbol{r}, \Psi)\right|, \left|t_l(\boldsymbol{r}, \Psi) - t_{l-1}(\boldsymbol{r}, \Psi)\right|} \tag{18}$$

where the index l describes successive loudspeakers within the array. Considering all control points and limitations of the microphone array, an upper frequency for loudspeaker l is derived,

$$f_{\mathrm{al}}^{\mathrm{meq}}(\mathrm{ls}_l) = \min[f_{\mathrm{al}}^{\mathrm{mic}}, \min_{j=1,\ldots,L} f_{\mathrm{al}}^{\mathrm{ls}}(\mathrm{ls}_l, \boldsymbol{r}_j, \Psi)]. \tag{19}$$

It defines the upper frequency for the calculation of filters from multichannel equalization [6]. Subsampling is then applied to both A and $\tilde{C}$ in order to reduce their lengths and improve filter calculation speed.

3.4 Above Aliasing Frequency

Since no accurate control can be achieved above the aliasing frequency, individual equalization is used (see the Introduction). Filters are derived from $C(z)$ using measurements corresponding to microphones located in a given solid angle around the axis of a loudspeaker (see Fig. 12). The current method considers a solid angle of 60° around the main axis. The magnitude of selected impulse responses is smoothed using a nonlinear method that preserves peaks and compensates for dips and is similar to the one presented in [16]. The responses obtained are averaged in the frequency domain and inverted. Individual equalization filters are then synthesized as minimum-phase filters.

Complete filters are composed by applying a delay and a gain to the individual equalization filters to achieve a smooth transition with filters obtained from multichannel inversion. Thanks to the modified multichannel scheme, delays are τ_{eq} for all loudspeakers and any considered virtual source (see Section 1.2). Gains to be applied to the filters are obtained from WFS driving functions [see Eq. (10)].

A final step consists of estimating the response at the microphone position for each source by processing the measured impulse responses with the filters obtained. Above the aliasing frequency the level is estimated in frequency bands and averaged over all microphones. Global correction gains are calculated in these frequency bands in order to ensure a "flat" mean frequency response [6], [16].

4 EVALUATION METHOD

In this section an evaluation scenario of the multichannel equalization is proposed. It is based on an estimate of the coloration introduced by the rendering system for the synthesis of an ensemble of virtual sources in an extended listening area.

4.1 Test Setup

The loudspeaker setup considered here is composed of 48 channels forming an 8.04-m-long linear array. This corresponds to a loudspeaker spacing of 167.5 mm. It is made of

- 48 ideal omnidirectional loudspeakers
- 6 eight-channel multi-actuator panels (MAPs) (see Fig. 13).

MAP loudspeakers have been recently proposed [45], [16], [18] as an alternative to electrodynamic loudspeakers for WFS. Using MAP loudspeakers, tens to hundreds of channels can easily be concealed in an existing environment given their low visual profile. They do, however, exhibit complex directivity characteristics, which have to be compensated for.

Measurements of MAP Loudspeakers System Measurements of the radiation of the MAP loudspeaker arrays have been carried out in the concert hall at IRCAM. Both loudspeakers and microphones were placed at 3.5 m from the floor. Their free-field contributions were extracted by windowing out the reflected contributions of the room.

A 24-channel linear microphone array (100-mm spacing) associated with a 24-channel preamplifier and a 24-

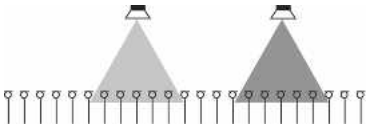

Fig. 12. Measurement selection for individual equalization.

channel sound card enabled simultaneous multichannel measurements of each loudspeaker channel. The microphone array was moved to four different locations to characterize the contribution of each loudspeaker channel on 96 regularly spaced (100-mm) positions at a given distance from the loudspeaker array. Four different depths (1.5, 2, 3, and 4.5 m from the loudspeaker array) were considered (see Fig. 14). The center of the chosen coordinate system is located at 3.5 m from the center of the loudspeaker array. All filters are designed such that all virtual sources are time and level aligned at this point. The microphone depths are referred to by their *Y* positions (2, 1.5, 0.5, and −1 m).

Test Sources A test ensemble of 15 omnidirectional virtual sources (see Fig. 14) is made of

- Five focused sources at 1 m (centered), 0.5 m, and 0.2 m (centered and off center) from the loudspeaker array (sources 1/2/3/4/5)
- Eight sources (centered and off center) behind the loudspeaker array at 0.2, 1, 3, and 8 m (sources 6/7/8/9/10/11/12)
- Two plane waves at 0° and 30° (sources 14/15).

The test ensemble chosen is a limited set of sources extracted from the grid proposed in the Introduction. It spans possible locations of virtual sources whose visibility area covers most of the listening space defined by the microphone arrays and represents "typical" WFS sources. In the proposed ensemble, some locations correspond to limit cases for WFS (focused sources, sources close to the loudspeaker array, sources at the limits of the visibility area).

Filter Calculation The proposed method (multichannel equalization, referred to as Meq in figures) is compared to a more classical method combining WFS driving functions and individual equalization (WFS + EQ). The individual equalization method is described in Section 3.4. Filter sets are calculated for both methods, the two loudspeaker types, and the ensemble of test sources.

Multichannel equalization is achieved by describing the MIMO system from the 96-channel measurements or simulations at $y = 1.5\ m$ (2 m from the loudspeaker array). Filters calculated for both individual and multichannel equalization are 800 taps long at a 48-kHz sample rate. For multichannel equalization a test scenario of the free parameters of the method ($d_{\mathrm{vis}}^{\mathrm{ls}}$, τ_{eq}, L_{filt}) is proposed.

4.2 Coloration Evaluation

Sound coloration introduced by a loudspeaker system is linked both to audible nonlinearities and to impairments in the linear behavior of the system. The latter is described by measuring an impulse response of the system at the listening position considered. Omitting nonlinearities, the flatness of the frequency response and the sharpness of the impulse response are usually used as objective measures of the transparency of the system at a given listening position. They affect the temporal and frequency-related aspects of the audio signal. In this part, only linear frequency-related impairments are considered. Temporal impairments are discussed in Section 6.

4.2.1 Sensitivity to Frequency Impairments

Frequency response impairments have been studied first by considering the audibility of resonances (and antiresonances) [46]–[48]. The audibility threshold is shown to vary depending on resonance level, resonance *Q* factor, and program material. Accounting for the most sensitive judgments, the audibility level threshold for resonances (and antiresonances) may be set to ±1 dB.

Moore and Tan [27] recently proposed an objective criterion for coloration estimation introduced by elec-

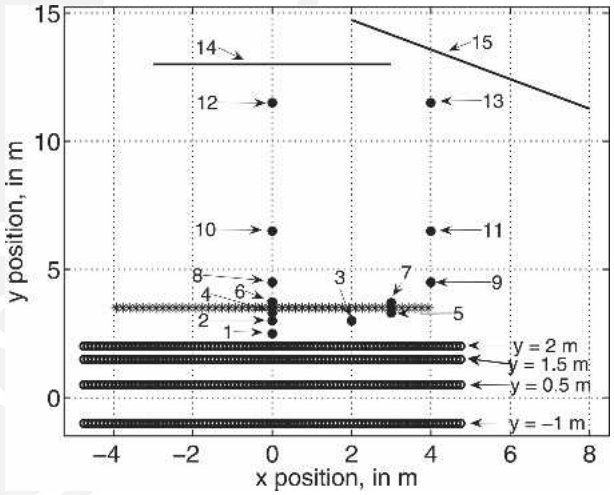

Fig. 14. Top view of system considered: 48 regularly spaced (167.5 mm) loudspeakers (*) measured on 4 depths ($y = -1$, 0.5, 1.5, 2.5 m) with 96 regularly spaced (100 mm) microphones (○) reproducing 15 test sources (●).

Fig. 13. MAP loudspeakers.

troacoustical systems. The model performs an analysis in the ERB_N (mean equivalent rectangular bandwidth of the auditory filters, N denoting normal hearing) frequency bands [49] and compares an original and a distorted representation of a signal. It extracts excitation levels for each ERB_N frequency band i of the original [EO(i)] and the distorted [ED(i)] signal. This measure enables one to account for the limited frequency resolution of the auditory system and its sensitivity to impairments in the frequency response of an electroacoustical system.

A first factor D_1 is calculated as the standard deviation of the first-order differences [$EO(i) - ED(i)$],

$$D_1 = \sigma\{W(i) \times [\mathrm{ED}(i) - \mathrm{EO}(i)]\} \tag{20}$$

where $W(i)$ are weights that are applied to account for the lower importance of some frequency bands in the estimation of coloration (below about 100 Hz and above 10 000 Hz). A second factor D_2 is calculated as the standard deviation of the second-order differences $\{[\mathrm{EO}(i+1) - \mathrm{ED}(i+1)] - [\mathrm{EO}(i) - \mathrm{ED}(i)]\}$,

$$D_2 = \sigma[W(i) \times \{[\mathrm{ED}(i+1) - \mathrm{ED}(i)] - [\mathrm{EO}(i+1) - \mathrm{EO}(i)]\}]. \tag{21}$$

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

$$D = w \times D_1 + (1 - w) \times D_2. \tag{22}$$

In [27] the weighting factor w is set to 0.4. D is shown to be linearly linked to subjective quality ratings for sound reproduction using headphones or loudspeakers in anechoic conditions [27]. This criterion may thus also be used for the evaluation of the free-field equalization of loudspeaker systems as presented in this paper. The D factor is not meant as an absolute judgment of the audibility of coloration artifacts. It is defined as a quality rating—the lower the D value, the better the quality.

4.2.2 Sound Coloration Evaluation for WFS

A complete sound coloration evaluation for WFS should account for its specificities:

- The listener is free to move in an extended listening area
- The system targets the synthesis of a large set of virtual sources, each having particular equalization filters
- Above the aliasing frequency, the frequency response is very much position dependent and cannot be evaluated at one spot only.

The first two specificities mentioned require that the sound coloration be evaluated for a large set of sources and listening positions. We propose to evaluate and compare the impulse response of the system $h_\Psi(\boldsymbol{r}_j, t)$ ($H_\Psi(\boldsymbol{r}_j, f)$ in the frequency domain with the ideal WFS response $A_\Psi(\boldsymbol{r}_j, t)$ [see Eq. (16)] for a listening position $\boldsymbol{r}_j$. A quality function $q_\Psi(\boldsymbol{r}_j, t)$ ($Q_\Psi(\boldsymbol{r}_j, f)$ in the frequency domain may be defined as

$$Q_\Psi(\boldsymbol{r}_j, f) = \frac{H_\Psi(\boldsymbol{r}_j, f)}{A_\Psi(\boldsymbol{r}_j, f)}. \tag{23}$$

This quality function enables one to compensate for level and propagation time at $\boldsymbol{r}_j$. Position- and source-dependent $D_\Psi(\boldsymbol{r}_j)$ can be extracted from $Q_\Psi(\boldsymbol{r}_j, f)$ by considering an original signal having white-noise characteristics.

Above the aliasing frequency the typical frequency response is very much position dependent. Dips in the frequency response at a given position may be compensated by peaks at similar frequencies at a 100-mm remote position [6]. A proper evaluation of the sound coloration introduced by WFS above the aliasing frequency should account for binaural decoloration [50]. Due to the complexity of the process description and since the proposed equalization method targets mainly the compensation of artifacts at lower frequencies, we propose to concentrate only on frequencies lower than the aliasing frequency. In the latter, $D_\Psi(\boldsymbol{r}_j)$ factors are thus calculated using only frequency bands whose center frequency is below the aliasing frequency for the virtual source Ψ and position $\boldsymbol{r}_j$.

5 RESULTS

Results are obtained by convolving the calculated filters with the measured impulse responses. In the following we therefore consider that the loudspeaker system is multilinear and that each loudspeaker is independent from the others. This statement could be verified for MAPs from intermodulation distortion measurements [16].

$D_\Psi(\boldsymbol{r}_j)$ values are calculated both for equalization methods and for all available measurements and sources. The calculation is achieved on 96 ERB_N bands for the entire audible frequency range. Using the ERB_N scale, center frequencies cf(i) are derived from ERB_N numbers i. For the simplification of calculations, ERB_N filtering is simply achieved by selecting frequency bins lying in $[cf(N - 0.5), \ldots, cf(N + 0.5)]$. For the calculation of $D_\Psi(\boldsymbol{r}_j)$ the corresponding energy is calculated. Due to available loudspeaker technical limitations, only ERB_N frequency bands above 150 Hz are considered for which sufficient energy can be produced. First both equalization methods are compared, showing mean values and 95% percentile values of $D_\Psi(\boldsymbol{r}_j)$ for various test scenarios.

5.1 Comparison with Individual Equalization

The comparison against individual equalization is achieved for

- Each measurement distance for all sources
- Each source for all measurements

For both scenarios the mean values and the 95% percentile values (threshold for which 95% of the data set has lower values) of $D_\Psi(\boldsymbol{r}_j)$ are determined. The first scenario shows how far the equalization process is accurate with regard to the distance from the loudspeaker array, particularly in the case of the multichannel equalization (Meq). The second establishes the virtual source dependency on the quality of the synthesized sound field. For both scenarios the multichannel equalization parameters are $L_{\text{filt}} = 800$ samples, $\tau_{\text{eq}} = 150$ samples, $d_{\text{vis}}^{\text{ls}} = 1.5$ m.

5.1.1 Measuring Distance Dependency

Fig. 15 shows the mean values and the 95% percentile values of the D factor depending on the measuring distance for both ideal and MAP loudspeakers. The indicated positions correspond to the Y coordinate of the microphone array (see Fig. 14).

For all measuring distances, multichannel equalization shows superior results compared to the combination of WFS and individual equalization considering only mean values. However, 95% percentile values show lower but more similar values. This is further explained next, considering each source separately.

Considering MAPs, the values are generally higher than for ideal loudspeakers and the quality degrades slightly with distance. This shows the limitations of the method. It cannot compensate for all artifacts of the rendering system at the control points and suffers from sound field description limitations due to the use of linear microphone arrays. It should be noted that the equalization quality at a measuring distance ($y = 2$ m) closer to the loudspeaker array than the control distance ($y = 1.5$ m) is still accurate. This is due to the reversibility of time in wave propagation, which is used, for example, for the synthesis of focused sources with WFS.

5.1.2 Source Dependency

Fig. 16 shows the mean values and the 95% percentile values of the D factor depending on the virtual source for both ideal and MAP loudspeakers considering all measuring distances. The source labels correspond to the one presented in Fig. 14. It can be seen that for all sources, the

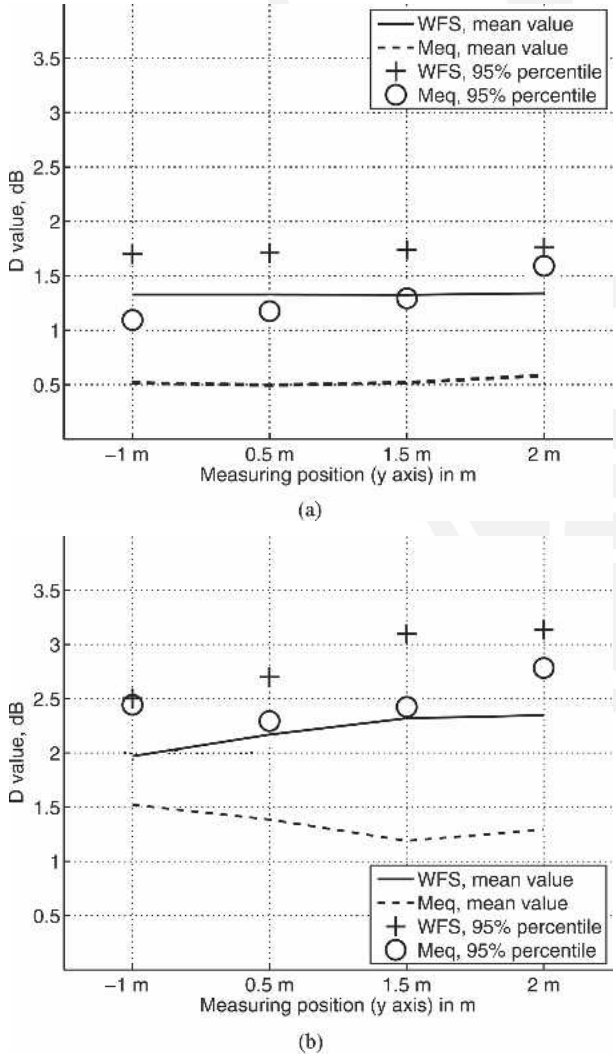

Fig. 15. Mean values and 95% percentile values of D factor dependency on measuring distance (see Fig. 14). $L_{\text{filt}} = 800$, $\tau_{\text{eq}} = 150$, $d_{\text{vis}}^{\text{ls}} = 1.5$ m. (a) Ideal loudspeakers, all sources. (b) MAPs, all sources.

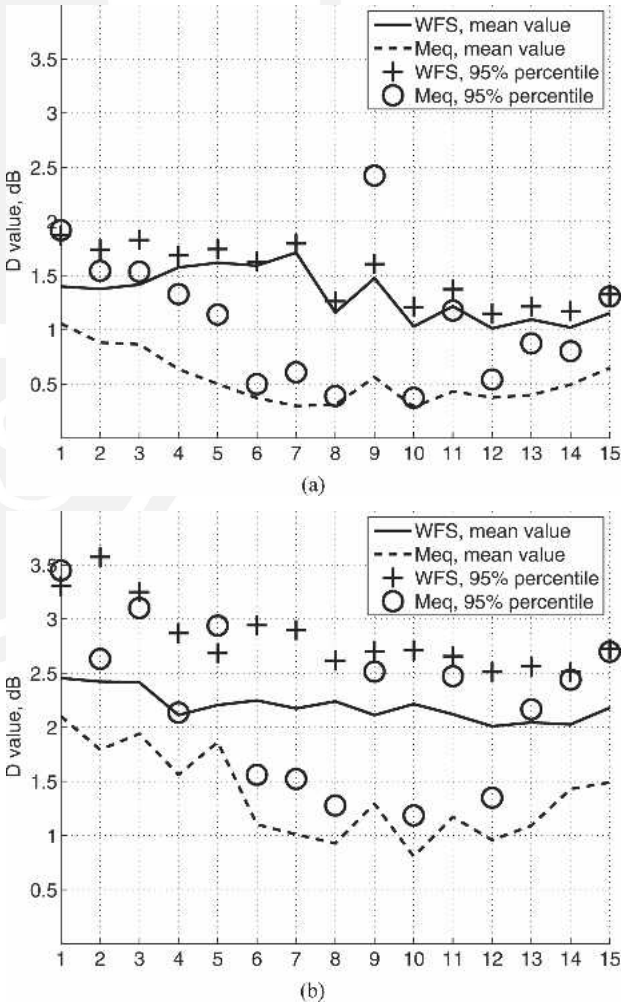

Fig. 16. Mean values and 95% percentile values of D factor dependency on virtual source (see Fig. 14). $L_{\text{filt}} = 800$, $\tau_{\text{eq}} = 150$, $d_{\text{vis}}^{\text{ls}} = 1.5$ m. (a) Ideal loudspeakers, all measuring distances. (b) MAPs, all measuring distances.

mean *D* factor value is lower for multichannel equalization than for the combination of WFS and individual equalization. However, 95% percentile values are obviously better for centered virtual sources (even numbers) but are of the same range for off-centered sources (odd numbers), especially for sources 9 and 11 (4 m to the right, 1 and 3 m behind the loudspeaker array). These are sources that suffer most from diffraction, which cannot be compensated for completely in the entire listening area. Fig. 17 shows *D* factor values for source 9 for all measuring positions. It can be seen that multichannel equalization focused more on listening positions directly facing the source (*x* positive positions) than on positions on the opposite side. At positions facing the source, the synthesized level of the source is higher than on the opposite side because the distance from the source is simply higher. The quality loss is thus explained, considering that the mean quadratic value used for error calculation is not normalized at each control point to the target level. Filters are thus optimized for positions where the target level of the sound field is higher.

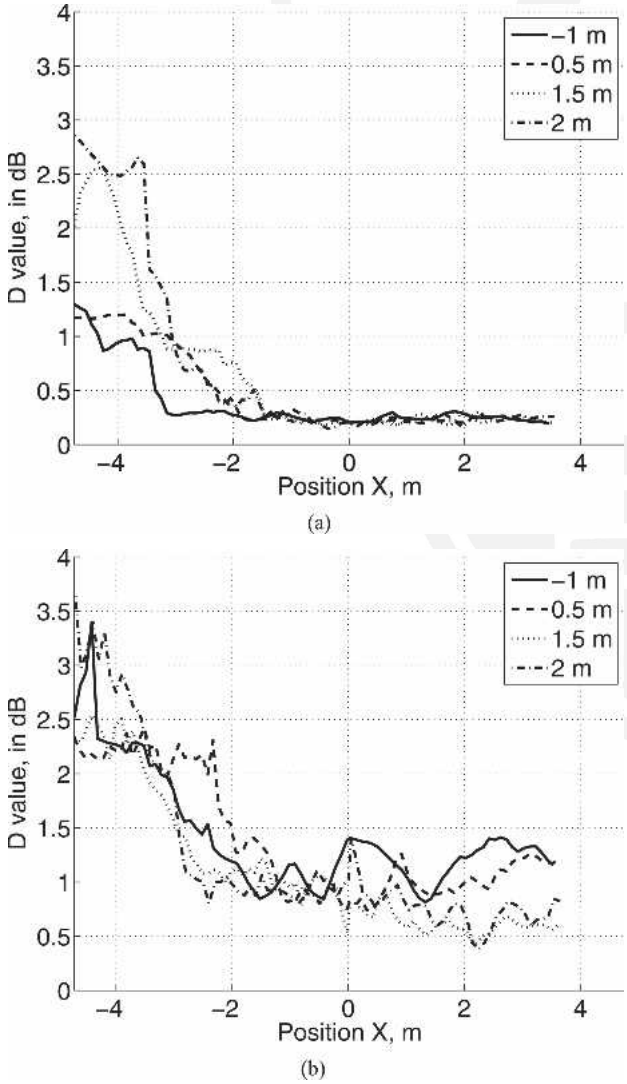

Fig. 17. *D* factor values calculated for multichannel equalization for source 9 for each measuring position (see Fig. 14). L_{filt} = 800, τ_{eq} = 150, d_{vis}^{ls} = 1.5 m. (a) Ideal loudspeakers, multichannel equalization. (b) MAPs, multichannel equalization.

5.2 Multichannel Equalization Method Optimization

This section proposes a parametric study for each optimization parameter of the method. Similar coloration evaluation is provided as in the previous part. Results are compared to the combination of WFS and individual equalization considering all sources and measuring distances.

5.2.1 Equalization Delay Dependency

Fig. 18 shows the mean values and the 95% percentile values of *D* factor dependency on the equalization delay τ_{eq} for all sources and measuring distances. Due to the limitations of backward masking [44], this predelay should be minimized to avoid the introduction of preechoes in the filters while maintaining a good equalization quality (see Section 3.3). Values for the two other parameters are L_{filt} = 800 and d_{vis}^{ls} = 1.5 m. Note that for individual equalization this value does not exist since minimum-phase filters are considered.

It can be seen that the value chosen (τ_{eq} = 150 samples, 3.1 ms at 48 kHz; see Section 4.1) is optimal for both ideal

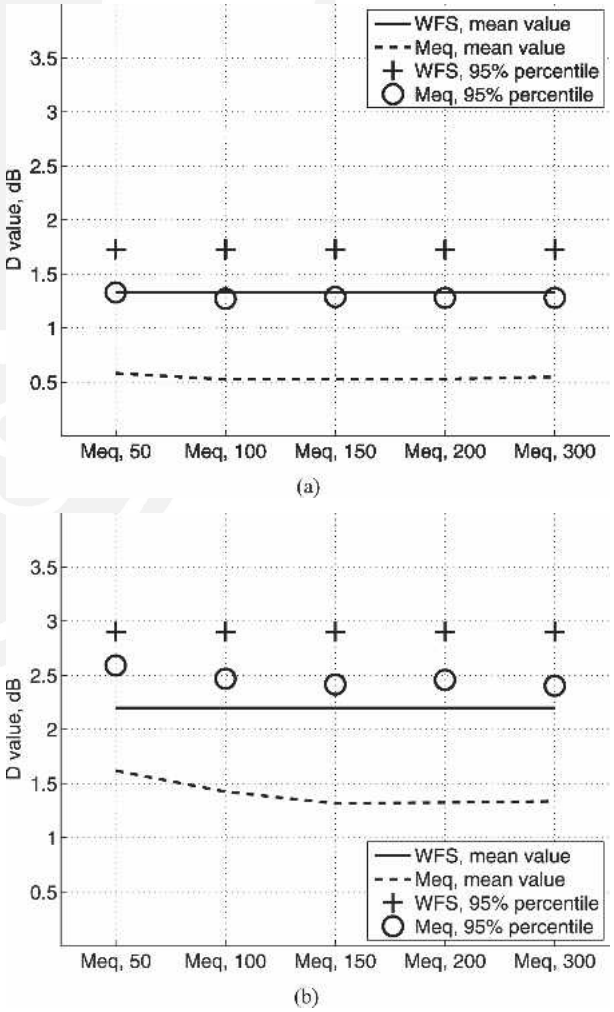

Fig. 18. Mean values and 95% percentile values of *D* factor dependency on equalization predelay τ_{eq} expressed as a number of sample periods at 48 kHz. L_{filt} = 800, d_{vis}^{ls} = 1.5 m. (a) Ideal loudspeakers. (b) MAPs.

and MAP loudspeakers. For ideal loudspeakers this value could even be reduced to 100 samples (2.1 ms at 48 kHz) without any significant effect. Both values are in the range of known limitations of backward masking (<5 ms at best) [44].

5.2.2 Filter Length Dependency

Fig. 19 shows the mean values and the 95% percentile values of D factor dependency on the filter length L_{filt} for all sources and measuring distances. This filter length should also be minimized. This limits the processing requirements for both filter calculation and real-time rendering. Values for the two other parameters are $\tau_{\text{eq}} = 150$ and $d_{\text{vis}}^{\text{ls}} = 1.5$ m. For comparison, individual equalization filters are 800 samples long.

Once again, the value chosen ($L_{\text{filt}} = 800$ samples) provides good results for both ideal and MAP loudspeakers. For ideal loudspeakers this value could even be reduced to 500 samples with very little effect. This holds true for the frequency range considered (150 Hz to aliasing frequency). For lower frequencies an additional analysis would be required.

5.2.3 Loudspeaker Selection Dependency

Fig. 20 shows the mean values and the 95% percentile values of D factor dependency on the loudspeaker selection parameter $d_{\text{vis}}^{\text{ls}}$ for all sources and measuring distances. This parameter determines the number of loudspeakers that need to be used for multichannel inversion. The analysis is reduced to sources 3/5/7/9. These are off-centered sources for which some loudspeakers may not be used for multichannel inversion according to the visibility criteria (see Section 3.3). In this case also, the minimization of $d_{\text{vis}}^{\text{ls}}$ enables one to improve the filter calculation speed and real-time processing by leaving out some loudspeakers. Values for the two other parameters are $\tau_{\text{eq}} = 150$ and $L_{\text{filt}} = 800$. The value chosen ($d_{\text{vis}}^{\text{ls}} = 1.5$ m) provides good results for both ideal and MAP loudspeakers. The use of all loudspeakers exhibits also low D values, but no real improvements over $d_{\text{vis}}^{\text{ls}} = 1.5$ m.

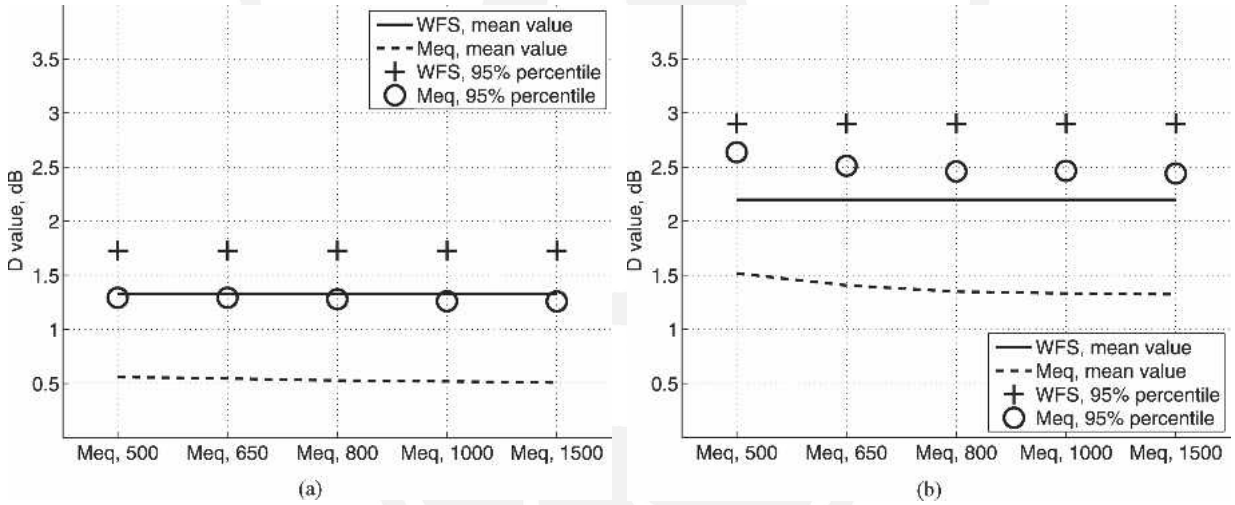

Fig. 19. Mean values and 95% percentile values of D factor dependency on filter length L_{filt}. $\tau_{\text{eq}} = 150$, $d_{\text{vis}}^{\text{ls}} = 1.5$ m. (a) Ideal loudspeakers, all sources and measuring distances. (b) MAPs, all sources and measuring distances.

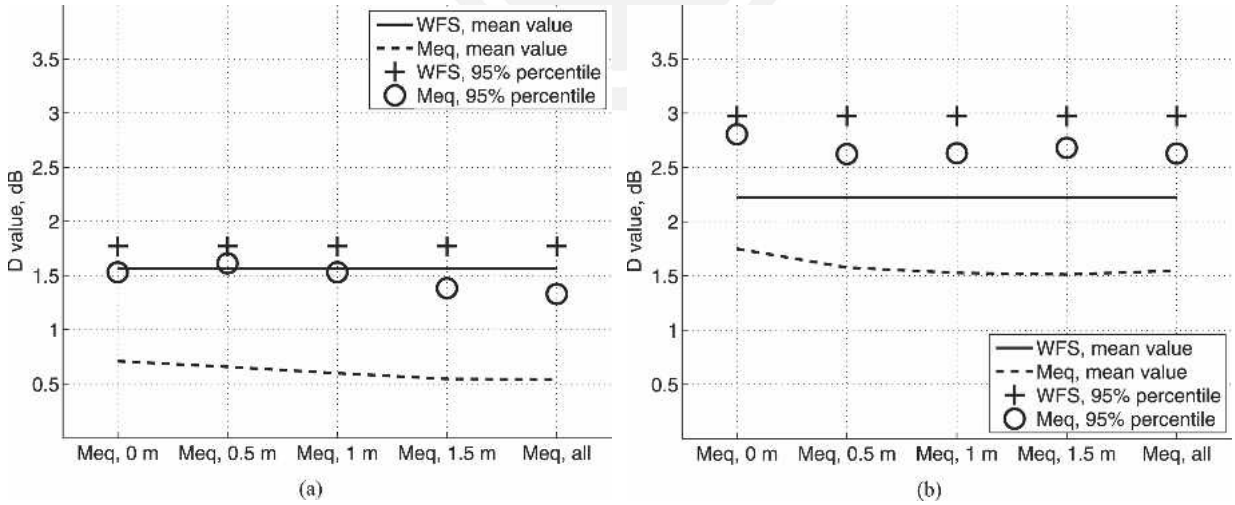

Fig. 20. Mean values and 95% percentile values of D factor dependency on loudspeaker selection $d_{\text{vis}}^{\text{ls}}$. $L_{\text{filt}} = 800$, $\tau_{\text{eq}} = 150$, only sources 3/5/7/9. (a) Ideal loudspeakers, all measuring distances. (b) MAPs, all measuring distances.

6 DISCUSSION AND PRACTICAL IMPLEMENTATION

6.1 Coloration Results

The objective criterion used in this paper is solely based on a perceptually motivated analysis of the frequency response. The principal results of the previous section are first summarized and discussed. The sensitivity to time-related impairments is then considered, outlining the lack of similar criteria.

6.1.1 Frequency-Related Coloration Evaluation

The free-field coloration evaluation proposed in Section 5 considered a test scenario for a linear 48-channel loudspeaker array and an ensemble of sources for two loudspeaker types—ideal omnidirectional loudspeakers and MAP loudspeakers. For both loudspeaker types the multichannel equalization method shows superior results to individual equalization and WFS. The D value is about 1 dB less on average (all measuring points, all sources), falling from 2.3 to 1.3 dB for MAP loudspeakers, and from 1.5 to 0.3 dB for ideal loudspeakers. Considering the curves provided in [27], which link the D factor to quality ratings, these are value ranges in which the lowering of the D factor corresponds to the steepest variations of the perceived quality for both music and speech. Therefore a significant quality improvement can be expected from the use of multichannel equalization. Only for off-centered sources does the quality degrade slightly for some positions located on the opposite side. These are, however, positions where the level is low and where the source may be partially masked while rendering a complete sound scene.

Improvements are obtained for all sources and all measuring distances, which proves the efficiency of the proposed method for sound-field equalization. The filter calculation is realized in most cases considering 48 loudspeakers, 96 microphones, and a filter length similar to the length of the measured impulse responses. This situation is therefore far from what the MINT theorem [12] requires (see Section 1.2) for complete multichannel inversion. However, the application considered does not target a complete inversion of the multichannel system, but rather a perceptually accurate control of the radiated sound field in an extended listening area.

6.1.2 Sensitivity to Temporal Impairments

Temporal impairments are analyzed from the phase or group delay response across frequency. Studies on the audibility of phase impairments consider simple all-pass filters to create group delay modifications in limited frequency bands [51]–[54]. They show that the audibility of temporal impairments is very dependent on the signal and listening conditions. For loudspeakers in reverberant conditions, impairments of less than 2 ms in a limited frequency band are usually considered inaudible. It is also usually accepted that in most electroacoustical systems, temporal-related impairments are inaudible or less audible than frequency-related impairments.

For an evaluation of the multichannel equalization method, impairments in the time domain are estimated by calculating the group delay in frequency bands $GD_\Psi[\boldsymbol{r}_j, ERB_N(i)]$ from $Q_\Psi(\boldsymbol{r}_j, f)$ using the same auditory filter bank as for the calculation of D. Results are reported in Table 1 for frequency bands below the aliasing frequency, considering all sources and listening positions. They show that multichannel equalization reduces both the mean values and the standard deviation of the group delay. However, levels are in the range of inaudibility when using the criteria mentioned before.

All studies mentioned considered group-delay impairments located after the main peak of the response. This is generally the case for electroacoustical systems such as electrodynamic or MAP loudspeakers. However, as soon as equalization is introduced, preringing may appear in the designed filters and more generally in the resulting impulse response. Considering the limitations of backward masking [44], the preringing may be audible. New coloration criteria involving time-domain impairment sensibility are thus required.

6.2 Real-time WFS Rendering

WFS rendering ideally relies on a content coding approach such as that defined in the MPEG4 standard [55]. In such an approach the sound scene is decomposed into an ensemble of virtual sources described by their positions, directivity characteristics, and virtual room effects (interaction with the virtual environment). A sound stream is attached to each source. Stereophonic encoded material could also be described in such a format by associating each channel with a "virtual" loudspeaker at the required position or direction.

For real-time rendering the WFS system is included in a rendering chain (see Fig. 21), which enables the synthesis of a synthetic room effect and authoring of the sound scene. The scene description parameters are transmitted in real time on a communication network [56]. The main linear loudspeaker array of the WFS system may be com-

Table 1. Mean value and standard deviation GD_{ERB} for all microphone positions and virtual sources.

	Mean Value		Standard Deviation	
Loudspeaker Type	WFS	Meq	WFS	Meq
Ideal	0.12 ms	0.01 ms	0.98 ms	0.31 ms
MAPs	1.07 ms	0.82 ms	1.43 ms	1.12 ms

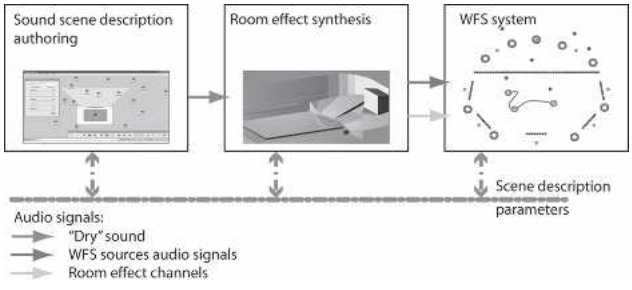

Fig. 21. WFS rendering chain.

pleted to the sides and rear with array portions or individual loudspeakers to render room effects or pan-potted sources [6], [57]. This type of system may be used for sound installations [58], concerts, or virtual and augmented reality applications.

Fig. 22 presents the architecture of a typical WFS rendering system. The complete system has a network structure. Rendering modules are associated with a group of loudspeakers. They get audio streams and "scene description" parameters from the communication network. Filters are calculated in an off-line process for the grid of target virtual sources. The calculation is achieved considering omnidirectional virtual sources and elementary directivity functions based on spherical harmonics [6], [59]. For each rendering engine, the filters corresponding to the associated loudspeakers are locally stored in a database. In the real-time mode, each rendering engine automatically loads filters associated with the closest source in the grid to the target source position and creates the required directivity characteristics as a combination of elementary directivity functions [59].

7 CONCLUSION

A multichannel equalization method is proposed that is dedicated to the equalization of sound fields radiated by loudspeaker arrays in an extended zone. The method uses a modified version of a multichannel inversion scheme. In such a scheme, the sound field radiated by the loudspeaker array is measured at a set of control points (spatially sampled) in order to create a MIMO system. Filters are thus calculated in reference to a target sound field (virtual source) which design the desired outputs of the MIMO system.

Unlike conventional multichannel inversion, the proposed method uses first the WFS framework to estimate the physical capabilities and limitations of the sound reproduction system. Delays necessary to wavefront forming are applied to the matrix of the impulse responses that describe the MIMO system. This creates a distinction between the reproduction of the target virtual source (multichannel sound reproduction) and the compensation of the rendering system deficiencies (equalization). The latter may be due to loudspeaker frequency response and directivity characteristics and limitations of WFS (see Section 2.2). The calculation of filters using multichannel inversion is limited to spatial and frequency regions where the loudspeaker array is capable of reproducing the target virtual source.

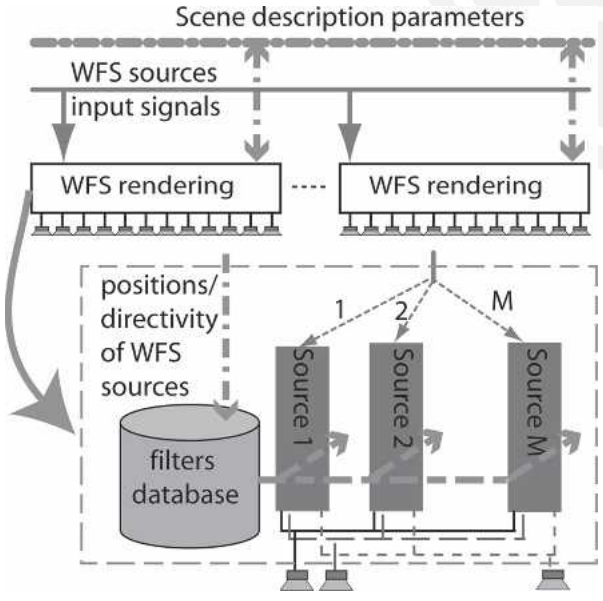

Fig. 22. WFS system rendering architecture.

Multichannel inversion usually allows for the control of a sound field produced by an ensemble of loudspeakers at a limited number of locations only. However, the control can be spatially extended by considering appropriate microphone array techniques. They provide a description of the sound field radiated by each loudspeaker that remains valid in an extended listening area. For compensation of linear loudspeaker array artifacts for WFS reproduction, it is proposed to rely on a linear microphone array that enables a description of the wave field radiated by the loudspeaker array in the horizontal plane.

The proposed method is compared in terms of coloration reduction to a more conventional equalization, which considers each loudspeaker separately. A parametric study was also provided to set the optimization parameters of the multichannel equalization method (filter length, number of loudspeakers considered equalization delay). The results show that the proposed method enables a partial but efficient compensation of rendering artifacts. This is particularly true in the case of MAP loudspeakers, which exhibit complex directivity characteristics. For ideal omnidirectional loudspeakers, the method also enables one to limit sound color variations due to inner limitations of WFS.

8 ACKNOWLEDGMENT

This work was partially funded by the European Carrouso project IST #1999-20993 and a grant from the French ministry of research. The author would like to thank Olivier Warusfel, Renato Pellegrini, Ulrich Horbach, Terence Caulkins, and Rik Van Zon for fruitful discussions on the topic.

9 REFERENCES

[1] A. J. Berkhout, "A Holographic Approach to Acoustic Control," *J. Audio Eng. Soc.,* vol. 36, pp. 977–995 (1988 Dec.).

[2] A. J. Berkhout, D. de Vries, and P. Vogel, "Acoustic Control by Wave Field Synthesis," *J. Acoust. Soc. Am.,* vol. 93, pp. 2764–2778 (1993).

[3] P. Vogel, "Application of Wave Field Synthesis in Room Acoustics," Ph.D. Thesis, Technical University Delft, Delft, The Netherlands (1993).

[4] E. W. Start, "Direct Sound Enhancement by Wave Field Synthesis," Ph.D. Thesis, Technical University Delft, Delft, The Netherlands (1997).

[5] R. Nicol, "Restitution Sonore Spatialisée sur une Zone Étendue: Application à la Téléprésence," Ph.D. Thesis, Université du Maine, Le Mans, France (1999); available at http://gyronymo.free.fr/audio3D/Guests/RozennNicol_PhD.html.

[6] E. Corteel, "Caractérisation et Extensions de la Wave Field Synthesis en Conditions Réelles d'Écoute," Ph.D. Thesis, Université de Paris VI, Paris, France (2004); available at http://mediatheque.ircam.fr/articles/textes/Corteel04a/.

[7] E. N. G. Verheijen, "Sound Reproduction by Wave Field Synthesis," Ph.D. Thesis, Technical University Delft, The Netherlands (1997).

[8] J. Blauert, *Spatial Hearing, The Psychophysics of Human Sound Localization* (MIT Press, Cambridge, MA, 1999).

[9] P. A. Nelson, F. Orduña Bustamente, and H. Hamada "Multichannel Signal Processing Techniques in the Reproduction of Sound," *J. Audio Eng. Soc.,* vol. 44, pp. 973–989 (1996 Nov.).

[10] D. H. Cooper and J. L. Bauck, "Prospects for Transaural Recording," *J. Audio Eng. Soc.,* vol. 37, pp. 3–19 (1989 Jan./Feb.).

[11] J. Bauck and D. H. Cooper, "Generalized Transaural Stereo and Applications," *J. Audio Eng. Soc.,* vol. 44, pp. 683–705 (1996 Sept.).

[12] M. Miyoshi and Y. Kaneda, "Inverse Filtering of Room Acoustics," *IEEE Trans. Acoust., Speech, Signal Process.,* vol. 36, pp. 145–152 (1988 Feb.).

[13] P. A. Nelson, H. Hamada, and S. J. Elliott, "Adaptative Inverse Filters for Stereophonic Sound Reproduction," *IEEE Trans. Signal Process.,* vol. 40, pp. 1621–1632 (1992 July).

[14] O. Kirkeby, P. A. Nelson, H. Hamada, and F. Orduña Bustamente, "Fast Deconvolution of Multichannel Systems Using Regularization," *IEEE Trans. Speech Audio Process.,* vol. 6, pp. 189–194 (1998).

[15] O. Kirkeby and P. A. Nelson, "Digital Filter Design for Inversion Problems in Sound Reproduction," *J. Audio Eng. Soc.,* vol. 47, pp. 583–595 (1999 July/Aug.).

[16] E. Corteel, U. Horbach, and R. S. Pellegrini, "Multichannel Inverse Filtering of Multiexciter Distributed Mode Loudspeaker for Wave Field Synthesis," presented at the 112th Convention of the Audio Engineering Society, *J. Audio Eng. Soc.* (*Abstracts*), vol. 50, p. 511 (2002 June), convention paper 5611.

[17] U. Horbach, E. Corteel, and D. de Vries, "Spatial Audio Reproduction Using Distributed Mode Loudspeaker Arrays," presented at the 21st International Conference of the Audio Engineering Society, Saint Petersburg, Russia (2002 June).

[18] M. M. Boone, "Multi-actuator Panels (maps) as Loudspeaker Arrays for Wave Field Synthesis," *J. Audio Eng. Soc.,* vol. 52, pp. 712–723 (2004 July/Aug.).

[19] T. Caulkins, E. Corteel, and O. Warusfel, "Wave Field Synthesis Interaction with the Listening Environment, Improvements in the Reproduction of Virtual Sources Located inside the Listening Room," presented at the 6th International Conference on Digital Audio Effects (DAFx-03), London, UK (2003 Sept.).

[20] O. Warusfel, E. Corteel, N. Misdariis, and T. Caulkins, "Reproduction of Sound Source Directivity for Future Audio Applications," presented at ICA04, Kyoto, Japan (2004 Apr.).

[21] E. Corteel and R. Nicol, "Listening Room Compensation for Wave Field Synthesis. What Can Be Done? presented at the 23rd International Conference of the Audio Engineering Society, Helsingør, Denmark (2003 June).

[22] R. van Zon, E. Corteel, D. de Vries, and O. Warusfel, "Multiactuator Panel (MAP) Loudspeakers: How to Compensate for Their Mutual Reflections?" presented at the 116th Convention of the Audio Engineering Society, *J. Audio Eng. Soc.* (*Abstracts*), vol. 52, p. 795 (2004 July/Aug.), convention paper 6052.

[23] S. Spors, A. Kuntz, and R. Rabenstein, "An Approach to Listening Room Compensation with Wave Field Synthesis," presented at the 24th International Conference of the Audio Engineering Society, Banff, BC, Canada (2003 June).

[24] S. Spors, H. Buchner, and R. Rabenstein, "Efficient Active Listening Room Compensation for Wave Field Synthesis," presented at the 116th Convention of the Audio Engineering Society, *J. Audio Eng. Soc.* (*Abstracts*), vol. 52, p. 812 (2004 July/Aug.), convention paper 6119.

[25] E. Hulsebos, D. de Vries, and E. Bourdillat, "Improved Microphone Array Configurations for Auralization of Sound Fields by Wave Field Synthesis," presented at the 110th Convention of the Audio Engineering Society, *J. Audio Eng. Soc.* (*Abstracts*), vol. 49, p. 532 (2001 June), convention paper 5337.

[26] S. Spors, M. Renk, and R. Rabenstein, "Limiting Effects of Active Room Compensation Using Wave Field Synthesis," presented at the 118th Convention of the Audio Engineering Society, *J. Audio Eng. Soc.* (*Abstracts*), vol. 53, p. 681 (2005 July/Aug.), convention paper 6400.

[27] 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," *J. Audio Eng. Soc.,* vol. 52, pp. 900–914 (2004 Sept.).

[28] T. D. Abhayapala, "Modal Analysis and Synthesis of Broadband Nearfield Beamforming Arrays," Ph.D. Thesis, Australian National University, Canberra, Australia (1999).

[29] S. Ise, "A Principle of Sound Field Control Based on the Kirchhof–Helmholtz Integral Equation and the Theory of Inverse System," *Acustica,* vol. 85, pp. 78–87 (1999).

[30] T. Betlehem and T. D. Abhayapala, "Theory and Design of Sound Field Reproduction in Reverberant Rooms," *J. Acoust. Soc. Am.,* vol. 117, pp. 2100–2111 (2005 Apr.).

[31] P. G. Craven and M. A. Gerzon, "Coincident Microphone Simulation Covering Three-Dimensional Space and Yielding Various Directional Outputs," U.S. patent 4,042,779 (1977).

[32] J. Daniel and S. Moreau, "Design Refinement of High Order Ambisonics Microphones—Experiments with a 4th-Order Prototype," presented at CFA/DAGA 04, Strasbourg, France (2004 Mar.).

[33] A. Laborie, R. Bruno, and S. Montoya, "A New Comprehensive Approach of Surround Sound Recording," presented at the 114th Convention of the Audio Engineering Society, *J. Audio Eng. Soc.* (*Abstracts*), vol. 51, p. 405 (2003 May), convention paper 5717.

[34] A. J. Berkhout, D. de Vries, and J. J. Sonke, "Array Technology for Acoustic Wave Field Analysis in Enclosures," *J. Acoust. Soc. Am.,* vol. 102, pp. 2757–2770 (1997).

[35] W. Putnam, D. Rocchesso, and J. O. Smith. "A Numerical Investigation of the Invertibility of Room

Transfer Functions," presented at the *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics,* Mohonk, NY (1995 Oct.).

[36] S. G. Norcross, G. A. Soulodre, and M. C. Lavoie, "Subjective Investigations of Inverse Filtering," *J. Audio Eng. Soc.,* vol. 52, pp. 1003–1028 (2004 Oct.).

[37] M. Bouchard, "Multichannel Affine and Fast Affine Projection Algorithms for Active Noise Control and Acoustic Equalization Systems," *IEEE Trans. Acoust., Speech, Signal Process.,* vol. 11, pp. 54–60 (2003 Jan.).

[38] P. A. Nelson, F. Orduña Bustamente, and H. Hamada, "Inverse Filter Design and Equalization Zones in Multichannel Sound Reproduction," *IEEE Trans. Speech Audio Process.,* vol. 3, pp. 185–192 (1995 May).

[39] J. C. Sarris, F. Jacobsen, and G. E. Cambourakis, "Sound Equalization in a Large Region of a Rectangular Enclosure," *J. Acoust. Soc. Am.,* vol. 116, pp. 3271–3274 (2004 Dec.).

[40] J. Sarris, N. Stefanakis, and G. Cambourakis, "Signal Processing Techniques for Robust Multichannel Sound Equalization," presented at the 116th Convention of the Audio Engineering Society, *J. Audio Eng. Soc.* (*Abstracts*), vol. 52, p. 805 (2004 July/Aug.), convention paper 2087.

[41] A. O. Santillan, "Spatially Extended Sound Equalization in Rectangular Rooms," *J. Acoust. Soc. Am.,* vol. 110, pp. 1989–1997 (2001 Oct.).

[42] M. Urban, C. Heil, and P. Bauman, "Wavefront Sculpture Technology," *J. Audio Eng. Soc.,* vol. 51, pp. 912–932 (2003 Oct.).

[43] E. Corteel, "On the Use of Irregularly Spaced Loudspeaker Arrays for Wave Field Synthesis, Potential Impact on Spatial Aliasing Frequency," presented at the 9th International Conference on Digital Audio Effects (DAFx-06), Montréal, PQ, Canada (2006 Sept.).

[44] T. G. Dolan and A. M. Jr. Small, "Frequency Effects in Backward Masking," *J. Acoust. Soc. Am.,* vol. 75, pp. 1508–1518 (1984 Mar.).

[45] M. M. Boone and W. P. J. de Bruijn, "On the Applicability of Distributed Mode Loudspeaker Panels for Wave Field Synthesis–Based Sound Reproduction," presented at the 108th Convention of the Audio Engineering Society, *J. Audio Eng. Soc.* (*Abstracts*), vol. 48, p. 364 (2000 Apr.), preprint 5165.

[46] R. Bücklein, "The Audibility of Frequency Response Irregularities," *J. Audio Eng. Soc.,* vol. 29, pp. 126–131 (1981 Mar.; transl. of article published in 1962).

[47] F. E. Toole and S. E. Olive, "The Modification of Timbre by Resonances: Perception and Measurement," *J. Audio Eng. Soc.,* vol. 36, pp. 122–142 (1988 Mar.).

[48] S. E. Olive, P. L. Schuck, J. G. Ryan, S. L. Sally, and M. E. Bonneville, "The Detection Thresholds of Resonances at Low Frequencies," *J. Audio Eng. Soc.,* vol. 45, pp. 116–128 (1997 Mar.).

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

[50] P. M. Zurek, "Measurements of Binaural Echo Suppression," *J. Acoust. Soc. Am.,* vol. 66, pp. 1750–1757 (1979).

[51] D. Preis, "Phase Distortion and Phase Equalization in Audio Signal Processing—A Tutorial Overview," *J. Audio Eng. Soc.,* vol. 30, pp. 774–794 (1982 Nov.).

[52] D. Preis and P. J. Bloom, "Perception of Phase Distortion in Anti-Alias Filters," *J. Audio Eng. Soc.,* vol. 32, pp. 842–848 (1984 Nov.).

[53] J. A. Deer, P. J. Bloom, and D. Preis, "Perception of Phase Distortion in All-Pass Filters," *J. Audio Eng. Soc.,* vol. 33, pp. 782–786 (1985 Oct.).

[54] S. Flanagan, B. C. J. Moore, and M. A. Stone, "Discrimination of Group Delay in Clicklike Signals Presented via Headphones and Loudspeakers," *J. Audio Eng. Soc.,* vol. 53, pp. 593–611 (2005 July/Aug.).

[55] R. Väänänen, O. Warusfel, and M. Emerit, "Encoding and Rendering of Perceptual Sound Scenes in the Carrouso Project," presented at the 22nd International Conference of the Audio Engineering Society, Espoo, Finland (2002 June).

[56] R. Pellegrini, M. Rosenthal, and C. Kuhn, "Wave Field Synthesis: Open System Architecture Using Distributed Processing," presented at the Forum Acusticum, Budapest, Hungary (2005 Sept.).

[57] M. Noguès, E. Corteel, and O. Warusfel, "Monitoring Distance Effect with Wave Field Synthesis," presented at the 6th International Conference on Digital Audio Effects (DAFX03), London, UK (2003 Sept.).

[58] G. Grand, E. Corteel, and R. Kronenber, "L'amiral cherche une maison à louer," Sound installation, Centre Georges Pompidou, DADA exhibition, 2005 Oct.–2006 Jan. (2005).

[59] E. Corteel, "Synthesis of Directional Sources Using Wave Field Synthesis, Possibilities, and Limitations," to be published, *J. Applied Signal Process.,* Special Issue Spatial Sound and Virtual Acoustics (2007 Jan.).

THE AUTHOR

Etienne Corteel was born in Vernon, France, in 1978. He received an M.Sc. degree in telecommunication engineering in 2000, and a Ph.D. degree in acoustics and signal processing from Paris 6 University, Paris, France, in 2004.

He joined Studer Professional Audio AG, Regensdorf, Switzerland, in 2001, where he started to work on wave field synthesis in the context of the European Carrouso IST project #1999-20993. He followed up this research at IRCAM, Paris, France, between 2002 and 2004. Since 2005 he has shared his time between IRCAM and sonic emotion, Oberglatt, Switzerland. His research interests include the design and evaluation of spatial sound rendering techniques for virtual or augmented reality applications (sound installations, concerts, simulation environments) as well as spatial hearing and cross-modal interactions.

Dr. Corteel is an associate member of the Audio Engineering Society.