SPATIO-TEMPORAL SIGNAL PREPROCESSING FOR MULTICHANNEL ACOUSTIC ECHO CANCELLATION

Karim Helwani, Sascha Spors, and Herbert Buchner

Quality and Usability Lab, Deutsche Telekom Laboratories, Technische Universität Berlin
Ernst-Reuter-Platz 7, 10587 Berlin, Germany
Email:{karim.helwani, sascha.spors}@telekom.de, hb@buchner-net.com

ABSTRACT

Hands-free full-duplex communication systems require acoustic echo cancelers to reduce echoes. Multichannel sound reproduction enhances realism in virtual reality and multimedia communication systems. However, in the case of multichannel systems the acoustic echo cancellation problem is challenging because of the non-uniqueness of the solution in the least squares sense. Therefore, preprocessing techniques are required. Known preprocessing approaches act in the temporal domain. In this paper we propose an acoustic echo cancellation preprocessing stage which bases on the spatial diversity offered by massive multichannel reproduction systems.

Index Terms— Array signal processing, Echo cancellers, Spatial filters.

1. INTRODUCTION

Whenever hands-free and full-duplex communication is desired, acoustic echo cancellation (AEC) is required. A typical scenario for stereo or multichannel AEC can be described as follows: From a transmission room, a sound source (e.g., a speaker) is picked up by P microphones ($P = 2$ for stereo). The microphone signals are transmitted to a receiving room and reproduced via P loudspeakers. At the same time, Q microphones in the receiving room pick up speech from a local user. In order to prevent the sound emitted from the loudspeakers coupling into the outgoing microphone signal (which is sent back to the far-end listener or some multimedia terminal), AEC attempts to cancel out any contributions of the incoming loudspeaker signals from the microphone signal by subtracting filtered versions of the loudspeaker signals from the microphone signal. AEC is a system identification problem that can be solved by adaptive realizations of Wiener filtering [1]. Wiener filters are the optimal solution in a linear least squares error (LSE) sense. In practice, adaptive filters are used to cope for time-varying systems. The solution of the adaptive filters converges asymptotically (in the mean) to the Wiener solution. It can be shown that the speed of the convergence depends on the eigenvalue spread or the ratio of maximum to minimum eigenvalue of the data autocorrelation matrix [1]. A large ratio is linked to a poor system excitation, such that some of the eigenmodes are not stimulated.
It has been shown that cross-correlations between the loudspeaker signals let the adaptive filter converge to a solution that depends on the characteristics of the loudspeaker signals. Any movement of the sound source in the transmission room results in a breakdown of the echo cancellation performance and requires a new adaptation of the cancellation filters [2]. Therefore, a preprocessing stage to decorrelate the transmitted signals for a unique identifiability of the echo paths is required to ensure robustness to sound source movements.
High resolution spatial sound field synthesis techniques synthesize a desired wave field within a given listening area. Closely spaced arrays of a large number (tens to hundreds) of individually driven loudspeakers are required for a physically accurate synthesis. Well known techniques are Wave Field Synthesis (WFS) [3, 4] and Higher-Order Ambisonics (HOA) [5].
Typically, the driving signals of the multichannel reproduction system are not only auto-correlated but also highly cross-correlated. This results in an ill-conditioned correlation matrix in the underlying normal equation of the MIMO adaptive filter [6]. In addition, the WFS technique performs loudspeaker selection [4], i.e., certain loudspeakers are active for some source positions and inactive for other positions. Hence, the ill-conditioning of the normal equation is not only the result of exciting the system with correlated inputs but it also emerges from the fact that certain inputs of the system might be set to zero depending on the source position. This motivates the need for a preprocessing stage to uniquely identify such multichannel systems. Key requirements for preprocessing schemes are not only convergence enhancement and complexity, but also subjective sound quality, i.e., such a preprocessing unit must not introduce any objectionable artifacts into the reproduced audio signals.

1.1. State of the art: Temporal preprocessing

A first simple preprocessing method for stereo AEC was proposed in [2] and achieves signal decorrelation by adding nonlinear distortions to the signals. While this approach features extremely low complexity, the introduced distortion products can become quite audible and objectionable, especially for high-quality applications using music signals. Moreover, the generalization of this approach to an arbitrary number of channels is not straightforward.
A second well-known approach consists of adding uncorrelated noise to the signals. In [7], this is achieved by introducing uncorrelated quantization distortion that is masked according to a psychoacoustic model. In principle, this approach is able to prevent audible distortions for arbitrary types of audio signals and may be generalized to more than two channels. However, the associated complexity and the introduced delay render this approach unattractive for most applications.
A perceptually well motivated preprocessing approach based on a frequency-selective phase modulation below the threshold of human perception was presented in [8]. This approach has been demonstrated to be effective in 5-channel surround sound echo cancellation in combination with a fast multichannel adaptation algorithm and is suitable for the generalization to the multichannel case. Here, the input audio signal is decomposed into subband signals by means of an analysis filterbank. The subband phases are modified based on a

set of frequency-dependent modulating signals.
Unfortunately, preliminary experiments of the authors have shown that *massive* multichannel reproduction systems, e.g., WFS, are very sensitive to phase modulations on the driving signals, especially at low frequencies. However, the idea of randomizing the phase of the driving function above the aliasing frequency of the system is not uncommon in the practical implementations of wave field synthesis systems, e.g., in [9]. The basic idea of this approach is to smear out the spatial structure of spatial aliasing in terms of enhancing the auditory event above the the spatial aliasing frequency. Another approach presented in [10] using diffuse filters has also shown some potential in reducing the coloration problems of WFS systems above its aliasing frequency.

1.2. Novel concept: Spatio-temporal preprocessing

Fortunately, multichannel systems entail possibilities to cope with the mentioned problems, e.g., in [11] the directivity control, offered by microphone arrays, was exploited aiming at suppressing the short-range acoustic feedback from the loudspeakers to the array output, hence resulting in lower acoustic echo. In this paper we show how multichannel acoustic *reproduction* systems can make a significant contribution to reduce the acoustic echo.
As mentioned, practical implementations of WFS and HOA use densely spaced loudspeaker arrays. The loudspeaker distance limits the frequency up to which the sound field can be controlled [12]. Typical setups have a loudspeaker spacing of $10 \ldots 30$ cm which allows to control the wave field up to approximately $1 \ldots 2$ kHz. This controllability of the sound field for low frequencies on the one hand, and the tight psychoacoustic restrictions as mentioned above on the other hand motivate the novel frequency-selective spatio-temporal preprocessing outlined in this paper.
Phase modulation should take place only over the frequency up to whom the system re-syntheses trusty spatial information. The preprocessed signal enhances the conditioning of the identification problem, and the acoustic system between the active loudspeakers and the recording system can be identified by known approaches for multichannel acoustic echo cancellation, such as proposed in [6, 13].

2. SYNTHESIS OF ACOUSTIC WAVE FIELDS USING LINEAR LOUDSPEAKER ARRAYS

Acoustic multichannel reproduction systems aim at synthesizing a desired wave field in a certain region by driving loudspeakers (secondary sources) on the boundary of this region. In [14] a method, called Spectral Division Method (SDM), was introduced to obtain analytically the driving functions. It bases on the idea that the synthesis can be understood as convolution and the synthesis equation for linear loudspeaker distributions reads [1]

$$P(\mathbf{x},\omega) = \int_{-\infty}^{\infty} D(\mathbf{x_0},\omega) G(\mathbf{x}-\mathbf{x_0},\omega) d\mathbf{x} \tag{1}$$

thereby, the secondary source distribution is assumed to be along the x-axis thus $\mathbf{x}_0 = [x_0, 0, 0]$ and $\mathbf{x} = [x, y_{\mathrm{ref}}, 0]$ defines a reference line on which the reproduction should be perfect. The secondary sources are driven by the signal $D(\mathbf{x}_0, \omega)$. $G(\mathbf{x}-\mathbf{x}_0, \omega)$ denotes the spatio-temporal transfer function of the secondary source located at $\mathbf{x}_0$, i.e., the temporal spectrum of the sound field it emits when it is fed by a temporal impulse, and ω is the radial frequency. Note that $G(\cdot)$ is assumed to be shift invariant.
We give here exemplarily the explicit formula for the driving signals to reproduce a virtual plane wave of given propagation direction and frequency. Since complex sound fields can be represented by plane waves via the angular spectrum representation [15] , the obtained results can be straightforwardly generalized.
For convenience, we want to reproduce a virtual plane wave which propagates along the x-y-plane

$$p(x, y, t) = e^{-j(k_{\mathrm{pw},x} + k_{\mathrm{pw},y} - \omega t)}, \tag{2}$$

with $[k_{\mathrm{pw},x} \quad k_{\mathrm{pw},y}] = k_{\mathrm{pw}}[\cos(\theta_{\mathrm{pw}}) \quad \sin(\theta_{\mathrm{pw}})]$, θ_{pw} denotes the propagation direction of the plane wave in the x-y-plane, $k = \frac{\omega}{c}$, c is the speed of sound, and $j^2 := -1$. Performing a Fourier transformation w.r.t the time and then along the x-axis leads to the equation

$$\tilde{P}(k_x, y, \omega) = 4\pi^2 \delta(k_x - k_{\mathrm{pw},x}) \delta(\omega - \omega_{\mathrm{pw}}) e^{-j(k_y - k_{\mathrm{pw},y})y}. \tag{3}$$

Under the assumption that $\tilde{G}(.)$ does not exhibit zeros in the direction of the angle of incidence of the plane wave, the driving function can explicitly be computed by

$$\tilde{D}(k_x, y, \omega) = \frac{\tilde{P}(k_x, y, \omega)}{\tilde{G}(k_x, y, \omega)}. \tag{4}$$

Performing an inverse Fourier transform with respect to k_x on (4) yields the driving function $D(x, y, \omega)$ in temporal spectral domain. For practical implementations the driving function has to be sampled at the positions of the loudspeakers.

3. SPATIAL SELECTIVITY OF SYNTHESIZED SOUND FIELDS

Synthesizing acoustic wave fields with loudspeaker arrays offers the ability to perform spatial selection in the listening area. The selectivity can be realized either as beamforming, or as space division.

3.1. Beamforming with loudspeaker arrays

Since real-life applications of beamforming and null-steering techniques in microphone arrays are straightforward, many studies on designing such beamformers can be found [16, 17]. Due to the reciprocity principle of acoustics, the paradigm of microphone array technology can be reversed [18]. Since null-steering and beamforming techniques take only the angular distribution of the field energy into account, they can be emulated by WFS or HOA by neglecting or synthesizing only plane waves from specific directions.
Performing null-steering on the reproduction side in the direction in which the microphone array is pointing will ideally result in cutting off the feedback channel of the full-duplex communication system. However, most beamforming techniques usually make an implicit far field assumption, therefore, they are not suitable for the reproduction of arbitrary wave fields. Moreover, the generalization of the beamforming techniques for two dimensional array geometries enclosing the listening area is not straightforward.

3.2. Space division

A goal of some recent work was achieving spatial selectivity of a synthesized sound field by defining closed regions of quiet in the listening area. E.g., the technique of acoustic contrast control [19]

[1] Upper case denotes the temporal frequency domain. The spatial frequency domain (wavenumber domain) is indicated by a tilde over the respective symbol

addresses maximum brightness in a zone and maximizing the contrast between the bright zone and the quiet zone. This approach aims at finding an optimal solution with respect to the energy and does not care about the spatial information in the desired wave field. Moreover, since the inputs of this optimization approach are quantities on selected points in the listening area, the optimization will converge to local optimal solutions. In [20] an approach for creating zones of quiet with circular arrays is described. The authors propose using higher order spatial harmonics to cancel the undesirable effects of the lower order harmonics of the desired soundfield on the zone of quiet. This approach can not be applied for arrays with other geometries than circular ones.

3.3. Analytical approach for achieving spatial diversity with linear loudspeaker array

Reproducing zones of quiet nearby a desired acoustic wave field with linear secondary source distributions can be achieved by multiplying the desired wave field on the line, on which the reproduction is correct, (see Sec. 2), with a rectangular window with width $(1/a)$, denoted by $\Pi_x(a\mathbf{x})$.
To derive the driving functions of the loudspeakers according to the method introduced in Sec. 2, we need to transform the window into the k_x-space

$$\Pi_x(a\mathbf{x}) \circ\!\!-\!\!\bullet \frac{1}{\sqrt{2\cdot\pi\cdot a^2}}\,\mathrm{sinc}\left(\frac{k_x}{2\cdot\pi\cdot a}\right). \tag{5}$$

The windowed field $P_\Pi(x,y,\omega)$ in the k_x-space is given by convolving the desired plane wave with the transformed window function

$$\tilde{P_\Pi}(k_x,y,\omega) = \frac{1}{a}\tilde{\Pi}_x\left(\frac{k_x}{a}\right) * \tilde{P}(k_x,y,\omega) \tag{6}$$

$$= \frac{4\pi^2}{\sqrt{2\cdot\pi\cdot a^2}}\,\mathrm{sinc}\left(\frac{k_x-k_{\mathrm{pw},x}}{2\cdot\pi\cdot a}\right)\delta(\omega-\omega_{\mathrm{pw}})e^{-j(k_y-k_{\mathrm{pw},y})y}. \tag{7}$$

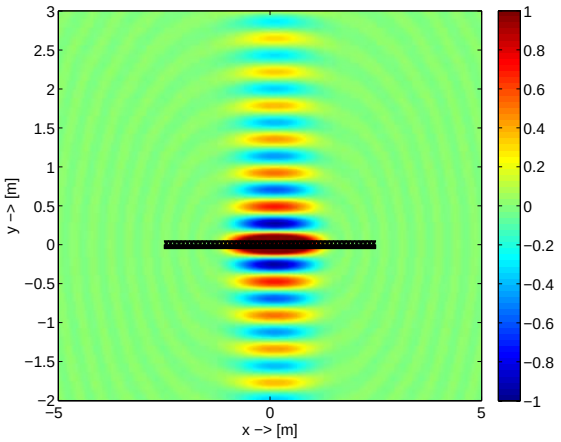
Hence, the driving function is given by

$$\tilde{D}(k_x,y,\omega) = \frac{\tilde{P_\Pi}(k_x,y,\omega)}{\tilde{G}(k_x,y,\omega)}. \tag{8}$$

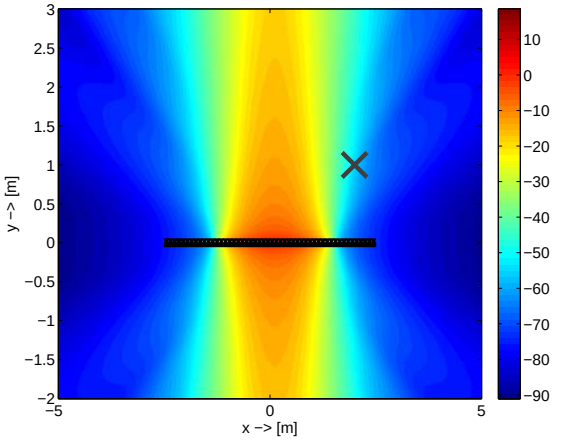
Due to the finite length of the array in practical implementations, other window functions with faster side-lobe decay are preferred, e.g., Hann or Blackman window.
The generalization of this approach to rectangular geometries can be achieved by taking the y-axis into account and multiplying the desired wave field with a two dimensional window instead of one dimensional ones, (see Fig. 3).

4. EXPERIMENTS

To illustrate the theoretical derivations outlined above, we present simulations of a sample scenario. We simulated a linear distribution of 50 omnidirectional secondary sources, separated by 10 cm, the window chosen was Hann-window and, the desired sound field is a plane wave, whose angle of incidence is $\pi/2$ to the x-axis. The chosen frequency was 800 Hz. In Fig. 1(a) the real part of the syntesized wave field is depicted. Fig. 1(b) shows the energy distribution of the synthesized wave field. To show the frequency dependent performance of the represented approach we computed the energy of the acoustic field at a point outside the desired wave field at the point $\mathbf{x} = [2\mathrm{m}, 1\mathrm{m}, 0]$ over the frequencies $20\ldots2500$ Hz. The result is



(a) real part of the synthesized acoustic wave field.



(b) energy distribution of the synthesized acoustic wave field.

Fig. 1. Synthesized wave field with a linear array of 50 loudspeakers at 800 Hz

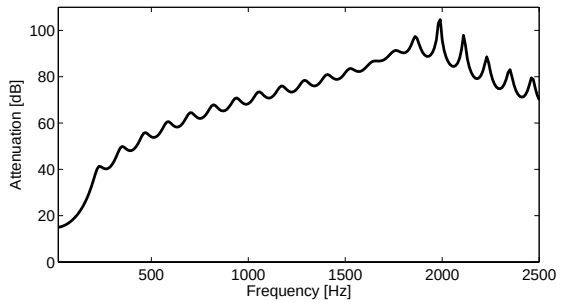
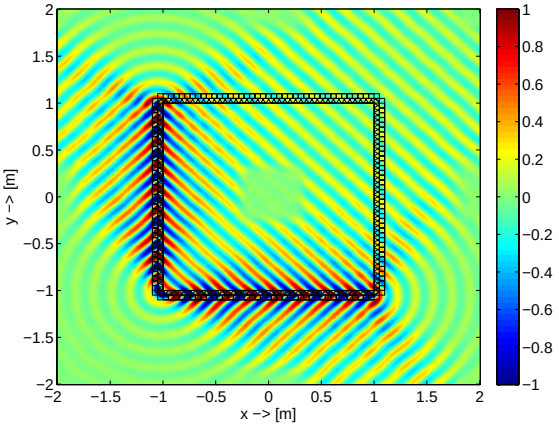


Fig. 2. Overall reached attenuation of the wave field synthesized by a linear array of 50 loudspeakers at the microphone position $\mathbf{x} = [2\mathrm{m}, 1\mathrm{m}, 0]$, marked by $\times$ in Fig. 1(b).

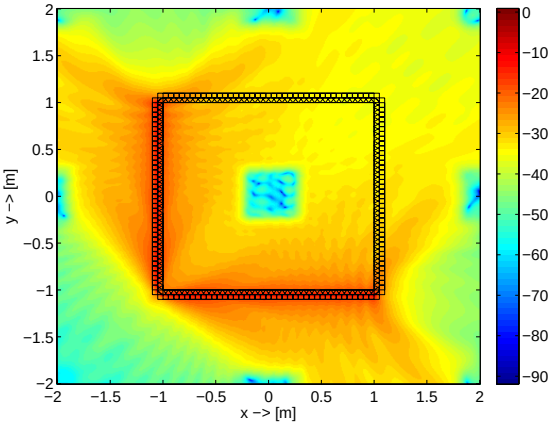
given in Fig. 2. The curve shows that the attenuation begins to fall down after passing the aliasing frequency of the system. However, the desired attenuation by the proposed preprocessing scheme covers a considerable range of the frequency band. Moreover, this result also motivates the integrated system with temporal processing above

the aliasing frequency, as outlined in Sec. 1.2

The generalization of the presented approach to rectangular arrays is illustrated by simulating a zone of quiet in the listening area of a rectangular speaker array. The dimension of the array was 2m×2m, the array is composed by 160 omnidirectional and equispaced loudspeakers. The desired wave field is a plane wave with a frequency of 2 kHz and $\pi/4$ as angle of incidence. The results are given in Fig. 3. In an AEC scenario the zone of quiet should be created in the area where the microphone array is usually placed, e.g., on a table in the middle of a teleconferencing room.



(a) real part of the synthesized wave field.



(b) energy distribution of the synthesized acoustic wave field.

Fig. 3. Synthesized wave field with a zone of quiet using a rectangular array of 160 loudspeakers. The desired sound field is a monochromatic plane wave of frequency 2000 Hz with unit amplitude and propagation direction $\theta_{\mathrm{pw}} = \pi/4$.

5. CONCLUSION

In this paper we have exploited the possibilities offered by loudspeaker arrays to overcome the problem of acoustic echo in full-duplex massive multichannel systems. We presented an analytical approach for creating zones of quiet with linear loudspeakers arrays that can be generalized for other geometries. The presented approach has two limitations emerging from the nature of sound field synthesis. The first limitation is the aliasing frequency of the loudspeaker array. The controllability of the sound field by known synthesis systems is available only up to this frequency. The second limitation is referred to the 2.5D problem. Thus, 2D wave field synthesis techniques aim at synthesizing two dimensional wave field with three dimensional point sources, this result in an amplitude decay of approximately 3dB per doubling of the distance when following the propagation path of a plane wave synthesized by the system [4, 14]. This limitation results in violating the orthogonality of the plane waves reproduced by such systems. Therefore, the construction of the zones of quiet has to deal with this limitation. These mentioned problems limit implicitly also the techniques of beamforming, null-steering, and acoustic contrast control. The presented approach offers therefore, insights into the limitations given by the physics of the spatial selectivity in synthesized acoustic wave fields. Moreover, the effort required for solving the acoustic echo cancellation problem can be done in a distributed manner, since our approach lets the reproduction system make a contribution in the cancellation process.

6. REFERENCES

[1] S. Haykin, *Adaptive filter theory*, Prentice Hall, Inc., 1991.

[2] J. Benesty, D.R. Morgan, and M.M. Sondhi, "A better understanding and an improved solution to the specific problems of stereophonic acoustic echo cancellation," *Speech and Audio Processing, IEEE Transactions on*, vol. 6, no. 2, pp. 156–165, 1998.

[3] A.J. Berkhout, D. De Vries, and P. Vogel, "Acoustic control by wave field synthesis," *The Journal of the Acoustical Society of America*, vol. 93, pp. 2764–2778, 1993.

[4] S. Spors, R. Rabenstein, and J. Ahrens, "The theory of wave field synthesis revisited," *Audio Eng. Soc. Conv. Paper*, vol. 124, 2008.

[5] J. Daniel, *Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia*, Ph.D. thesis, Université Paris 6, 2000.

[6] S. Spors, H. Buchner, and K. Helwani, "Block-based multichannel transform-domain adaptive filtering," in *European Signal Processing Conference (EUSIPCO)*, Aug. 2009.

[7] T. Gänsler and P. Eneroth, "Influence of audio coding on stereophonic acoustic echo cancellation," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, May 1998, pp. 3649–3652.

[8] J. Herre, H. Buchner, and W. Kellermann, "Acoustic echo cancellation for surround sound using perceptually motivated convergence enhancement," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing ICASSP 2007*, Apr. 2007, vol. 1, pp. I–17I–20.

[9] E.W. Start, *Direct Sound Enhancement by Wave Field Synthesis*, Ph.D. thesis, Delft University of Technology, 1997.

[10] T. Caulkins, E. Corteel, K.V. NGuyen, R. Pellegrini, and O. Warusfel, "Objective and subjective comparison of electrodynamic and map loudspeakers for wave field synthesis," in *Audio Engineering Society Conference: 30th International Conference: Intelligent Audio Environments*, 3 2007.

[11] B. Debail and A. Gilloire, "Microphone array design with improved acoustic echo rejection," in *Proc. IEEE International Workshop on Acoustic Echo and Noise Control (IWAENC)*, Sept. 2001, pp. 55–58.

[12] S. Spors, "Spatial aliasing artifacts produced by linear loudspeaker arrays used for wave field synthesis," in *Second IEEE-EURASIP International Symposium on Control, Communications, and Signal Processing, Marrakech, Morocco*, 2006.

[13] H. Buchner, J. Benesty, and W. Kellermann, "Generalized multichannel frequency-domain adaptive filtering: efficient realization and application to hands-free speech communication," *Signal Processing*, vol. 85, no. 3, pp. 549–570, 2005.

[14] J. Ahrens and S. Spors, "Sound field reproduction using planar and linear arrays of loudspeakers," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 18, no. 8, pp. 2038–2050, 2010.

[15] E.G. Williams, *Fourier acoustics: sound radiation and nearfield acoustical holography*, Academic Press, 1999.

[16] M. Brandstein and D. Ward, *Microphone arrays: signal processing techniques and applications*, Birkhäuser, 2001.

[17] J. Benesty, *Microphone array signal processing*, Springer, Berlin, 2008.

[18] E. Mabande and W. Kellermann, "Towards superdirective beamforming with loudspeaker arrays," in *Conf. Rec. International Congress on Acoustics*, 2007.

[19] J.W. Choi and Y.H. Kim, "Generation of an acoustically bright zone with an illuminated region using multiple sources," *The Journal of the Acoustical Society of America*, vol. 111, pp. 1695–1700, 2002.

[20] T. Abhayapala and Y.J. Wu, "Spatial soundfield reproduction with zones of quiet," in *Audio Engineering Society Convention 127*, 2009.