EFFICIENT REALIZATION OF MODEL-BASED RENDERING FOR 2.5-DIMENSIONAL NEAR-FIELD COMPENSATED HIGHER ORDER AMBISONICS

Sascha Spors, Vincent Kuscher and Jens Ahrens

Quality and Usability Lab, Deutsche Telekom Laboratories, Technische Universität Berlin, Germany
{sascha.spors,jens.ahrens}@telekom.de

ABSTRACT

Near-field compensated higher order Ambisonics is a sound field synthesis technique which is based upon a mathematical representation in terms of surface spherical harmonics. The generation of loudspeaker driving signals using digital signal processing is a numerically challenging task due to the involved special functions. This paper presents an efficient algorithm for 2.5-dimensional near-field compensated higher order Ambisonics. It is based upon a parametric representation of recursive filters realized in first- and second-order sections. The performance of the proposed technique is evaluated at an illustrative example.

1. INTRODUCTION

Sound field synthesis (SFS) techniques aim at the synthesis of a desired sound field within an extended area using an ensemble of loudspeakers. It is assumed that the correct synthesis in the physical sense leads to improved perceptual properties in comparison with traditional approaches, like stereophony. An number of methods have been developed in the past decades. Well known analytic methods in this context are Wave Field Synthesis (WFS), near-field compensated higher order Ambisonics (NFC-HOA) and the Spectral Division Method (SDM). In SFS two different techniques can be distinguished for the rendering of virtual scenes: (i) data-based and (ii) model-based rendering. For data-based rendering a scene is recorded by sound field analysis techniques, for instance using spherical microphone arrays, and the loudspeaker driving signals are derived such that they synthesize the spatio-temporal structure of the recorded scene as closely as possible.

In model-based rendering, models of virtual sources are used which are then fed by the (dry) virtual source signal. Frequently used models are point sources or plane waves. A virtual scene is typically composed from multiple sources of both types. Most of the currently available implementations of NFC-HOA are data-based. In [1] an approach for 3D model-based NFC-HOA was proposed, 2D synthesis was briefly discussed. The present paper proposes an efficient and parametric algorithm for 2.5-dimensional NFC-HOA which is more practical than a 2D or 3D solution.

2. NEAR-FIELD COMPENSATED HIGHER ORDER AMBISONICS

The theory of NFC-HOA has originally been developed in [3] on basis of the traditional Ambisonics approach. The following section briefly reviews a generalized viewpoint as presented in [4]. We begin with a spatially continuous formulation of the problem that is later on sampled.

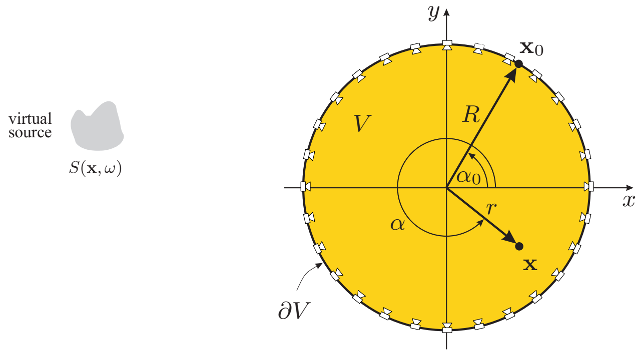

Figure 1: 2.5-dimensional synthesis of a virtual source using a circular distribution of secondary sources.

2.1. Basic principle

The theory of NFC-HOA is based on explicitly solving the underlying physical problem of synthesizing a desired virtual sound field within an extended area using an ensemble of loudspeakers (secondary sources). Analytic solutions are so far available for circular or spherical distributions of secondary sources. Here, we consider a circular distribution of secondary sources with monopole characteristics. The distribution does not enclose the listening volume, like for instance a spherical one would do. This constitutes a 2.5-dimensional scenario. Refer to [4] for a detailed discussion of the properties of 2.5-dimensional NFC-HOA.

Figure 1 illustrates the geometry underlying the following considerations. The sound field $P(\mathbf{x},\omega)$ synthesized by a circular distribution of secondary sources with radius R is given as

$$P(\mathbf{x},\omega) = \int_0^{2\pi} D(\alpha_0,\omega) G(\mathbf{x}-\mathbf{x}_0,\omega)\; R d\alpha_0 \;, \tag{1}$$

were $G(\mathbf{x}-\mathbf{x}_0,\omega)$ denotes the spatial transfer function of a secondary source at position $\mathbf{x}_0$, $D(\alpha_0,\omega)$ its weight (driving signal), $\mathbf{x} = r\,[\cos\alpha \;\; \sin\alpha]^T$ and $\mathbf{x}_0 = R\,[\cos\alpha_0 \;\; \sin\alpha_0]^T$. Expanding the quantities in (1) into a Fourier series with respect to the azimuth α yields

$$\mathring{P}_m(r,\omega) = 2\pi R\; \mathring{D}_m(R,\omega)\mathring{G}_m(r,\omega) \;, \tag{2}$$

were for instance $\mathring{P}_m(r,\omega)$ denotes the Fourier series coefficients of the synthesized sound field so that $P(\mathbf{x},\omega) = \sum_{m=-\infty}^{\infty} \mathring{P}_m(r,\omega) e^{jm\alpha}$. The secondary source driving signal in order to synthesize a sound field $S(\mathbf{x},\omega)$ can be determined to be

$$D(\alpha,\omega) = \frac{1}{2\pi R} \sum_{m=-\infty}^{\infty} \underbrace{\frac{\mathring{S}_m(r,\omega)}{\mathring{G}_m(r,\omega)}}_{\mathring{D}_m(\omega)} e^{jm\alpha} \;. \tag{3}$$

The undesired dependency on r of the Fourier series coefficients $\mathring{S}_m(r,\omega)$ and $\mathring{G}_m(r,\omega)$ in 2.5-dimensional synthesis can be overcome by referencing the synthesis to the center [4]. For simple source models, the required Fourier series coefficients can be derived analytically. In the following, we assume secondary monopole sources.

2.2. Spatial sampling of secondary source distribution

Equation (1) assumes a spatially continuous distribution of secondary sources. In practice, the distribution is realized by a spatially discrete distribution of individual loudspeakers. This constitutes a spatial sampling of the secondary source distribution. Refer for instance to [4, 5] for a discussion of spatial sampling in the context of NFC-HOA.

For a spatially discrete distribution of secondary sources, the Fourier series (3) is typically truncated. The truncation represents a spatial bandlimitation of the driving function. For an equiangular sampling of the secondary source distribution, the bandlimited driving signal is given as

$$D(\alpha,\omega) = \frac{1}{2\pi R} \sum_{m=-M}^{M} \mathring{D}_m(\omega) e^{jm\alpha} \;, \tag{4}$$

were $P = 2M + 1$ denotes the number of spatially discrete secondary sources (loudspeakers). For simplicity we limit our presentation to an odd number of loudspeakers P. The theory can be extended straightforwardly to even numbers by modifying the summation limits in (4).

2.3. Virtual plane waves and point sources

For model-based rendering, spatio-temporal models for the virtual source are required in order to derive the driving functions analytically. The most common models used are plane waves and point sources. We consider these two types in the remainder.

The Fourier series coefficients $\mathring{D}_{\text{pw},m}(\omega)$ of the driving function for the synthesis of a virtual plane wave $S_{\text{pw}}(\mathbf{x},\omega) = \hat{S}(\omega) e^{-j\frac{\omega}{c}\mathbf{n}_{\text{pw}}^T\mathbf{x}}$ using point sources as secondary sources are given as [4]

$$\mathring{D}_{\text{pw},m}(\omega) = \hat{S}(\omega) \underbrace{\frac{4\pi j(-j)^{|m|}}{\frac{\omega}{c}\, h^{(2)}_{|m|}(\frac{\omega}{c}R)}}_{H_{\text{pw,m}}(\omega)} e^{-jm\theta_{\text{pw}}} \;, \tag{5}$$

were $h^{(2)}_m(\cdot)$ denotes the m-th order spherical Hankel function of second kind, $\hat{S}(\omega)$ the time-frequency spectrum of the plane wave, θ_{pw} the azimuth of the propagation direction $\mathbf{n}_{\text{pw}} = [\cos\theta_{\text{pw}} \;\; \sin\theta_{\text{pw}}]^T$.

The Fourier series coefficients $\mathring{D}_{\text{ps},m}(\omega)$ of the driving function for a virtual point source $S_{\text{ps}} = \hat{S}(\omega) e^{-j\frac{\omega}{c}|\mathbf{x}-\mathbf{x}_{\text{ps}}|} / |\mathbf{x}-\mathbf{x}_{\text{ps}}|$ at position $\mathbf{x}_{\text{ps}} = r_{\text{ps}} [\cos\theta_{\text{ps}} \;\; \sin\theta_{\text{ps}}]^T$ are given as [4]

$$\mathring{D}_{\text{ps},m}(\omega) = \hat{S}(\omega) \underbrace{\frac{h^{(2)}_{|m|}(\frac{\omega}{c} r_{\text{ps}})}{h^{(2)}_{|m|}(\frac{\omega}{c}R)}}_{H_{\text{ps,m}}(\omega)} e^{-jm\theta_{\text{ps}}} \;. \tag{6}$$

Although some similarities exist compared to [1] our handling of 2.5-dimensional synthesis is different, leading for instance to $|m|$ in the formulas (5) and (6) above.

3. EFFICIENT REALIZATION OF NFC-HOA

We are seeking for a parametric and efficient realization of the 2.5D NFC-HOA driving functions for a plane wave and point source. The following subsections introduce the proposed approach.

3.1. Overall structure

According to (4), the driving function is given as a truncated Fourier series. For simplicity we assume an equiangular arrangement of P secondary sources at the angles $\alpha_{0,n} = {}^{2\pi n}/_{P}$. Introducing $\alpha_{0,n}$ into (4), reveals that the truncated Fourier series can be realized by an inverse Discrete Fourier Transformation or in practice very efficiently by an inverse Fast Fourier Transformation (IFFT) of length P. Due to the conjugate complex symmetry of the filter modes $H_{\text{pw},m}(\omega)$ or $H_{\text{ps},m}(\omega)$ and the exponential factors a complex to real valued IFFT may be used to further lower the computational complexity.

The Fourier series coefficients for a virtual plane wave (5) and a point source (6) are given by filtering the input signal $\hat{s}(t)$ for each mode m with a filter and multiplying the result with an exponential function. Figure 2 illustrates a block-diagram of the resulting overall signal processing structure. Note the real valued weight a and delay τ of the virtual source signal $\hat{s}(t)$ are introduced later. Since the filter modes depend only on the absolute value of m, it is sufficient to filter the input signal by $M + 1$ instead of $2M + 1$ filters. This, in conjunction with the IFFT, lowers the required computational complexity considerably. Note further that $H_0(\omega)$ is just a delay/weighting operation as is shown later.

In the following we focus on a parametric design of the filter modes. Since, the spherical Hankel function is a prominent part of the filter modes, we first review its realization as a recursive filter as presented in [1, 6].

3.2. Spherical Hankel function as recursive filter

In a first step, a series expansion of the spherical Hankel function is derived. Due to its close link to the z-transformation, it is useful to apply the Laplace transformation for the series representation. Using a series expansion of the spherical Hankel function and replacing $j\omega$ by s the desired expansion reads [6]

$$h^{(2)}_n(\tfrac{r}{c}s) = -j^n e^{-\frac{r}{c}s} \frac{\sum_{k=0}^{n} \beta_n(k) (\frac{r}{c})^k s^k}{(\frac{r}{c})^{n+1} s^{n+1}} \;. \tag{7}$$

The expansion coefficients $\beta_n(k)$ are given as

$$\beta_n(k) = \frac{(2n-k)!}{(n-k)!k!2^{n-k}} \;. \tag{8}$$

Note $\beta_n(k)$ can be calculated recursively by exploiting the recurrence relation of the spherical Hankel function [6]. A direct realization of (7) as digital recursive filter is likely to become numerically unstable for higher orders n. A decomposition into first- and second-order sections (FOS/SOS) is more stable in practice. Noting that the series expansion coefficients $\beta_n(k)$ are real-valued, Eq. (7)

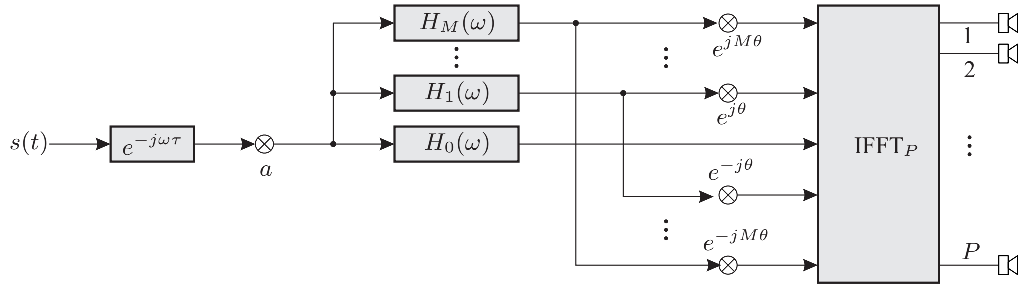

Figure 2: Block-diagram illustrating the signal flow of the efficient realization of model-based NFC-HOA.

can be factorized as follows

$$h_n^{(2)}(\tfrac{r}{c}s) = -j^n e^{-\frac{r}{c}s} \frac{1}{\frac{r}{c}s} \left(\frac{s - \frac{c}{r}\rho_0}{s} \right)^{\mathrm{mod}(n,2)} \times \prod_{l=1}^{\mathrm{div}(n,2)} \frac{(s - \frac{c}{r}\rho_l)^2 + \frac{c}{r}\sigma_l^2}{s^2} , \quad (9)$$

were ρ_0 denotes the real-valued root of the polynomial given by $\beta_n(k)$; ρ_l, σ_l denote the real and imaginary parts of its complex-conjugate roots, respectively. Equation (9) states that the roots of the denominator polynomial of (7) are given by scaling the roots of the normalized polynomial given by the coefficients $\beta_n(k)$. This is an important result for the desired parametric realization since only the roots of the normalized polynomial have to be computed. The next section illustrates how the series expansion (9) can be used to efficiently realize the filter modes $H_{\mathrm{pw},m}(\omega)$ or $H_{\mathrm{ps},m}(\omega)$ for model-based NFC-HOA.

3.3. Realization of filter modes

Introducing the FOS/SOS expansion (9) of the spherical Hankel function into the filter modes of a plane wave, as given by (5), yields

$$H_{\mathrm{pw},m}(s) = 4\pi R\, e^{\frac{R}{c}s} (-1)^{|m|} \left(\frac{s}{s - \frac{c}{R}\rho_0} \right)^{\mathrm{mod}(|m|,2)} \times \prod_{l=1}^{\mathrm{div}(|m|,2)} \frac{s^2}{(s - \frac{c}{R}\rho_l)^2 + \frac{c}{R}\sigma_l^2} . \quad (10)$$

In (10) $4\pi R$ represents a scaling and $e^{\frac{R}{c}s}$ an anticipation (negative delay) of the virtual source signal $\hat{s}(t)$ that can be discarded in practice. Otherwise causality has to be ensured by a pre-delay. The remaining terms can be realized by a digital filter consisting of FOS and SOS. Introducing (9) into the filter modes of a point source, as given by (6), yields

$$H_{\mathrm{ps},m}(s) = \frac{R}{r_{\mathrm{ps}}} e^{-\frac{r_{\mathrm{ps}}-R}{c}s} \left(\frac{s - \frac{c}{r_{\mathrm{ps}}}\rho_0}{s - \frac{c}{R}\rho_0} \right)^{\mathrm{mod}(|m|,2)} \times \prod_{l=1}^{\mathrm{div}(|m|,2)} \frac{(s - \frac{c}{r_{\mathrm{ps}}}\rho_l)^2 + \frac{c}{r_{\mathrm{ps}}}\sigma_l^2}{(s - \frac{c}{R}\rho_l)^2 + \frac{c}{R}\sigma_l^2} . \quad (11)$$

In (11) $\frac{R}{r_{\mathrm{ps}}}$ represents a scaling and $e^{-\frac{r_{\mathrm{ps}}-R}{c}s}$ a delay of the virtual source signal $\hat{s}(t)$. The scaling represents the amplitude decay of a point source, the delay its propagation time from the virtual source to the center. In both the dependency on R may be dropped in a practical implementation, or for certain applications it may even be desired to drop both the scaling and delay completely. Note the presented formulation allows to do so, while the scaling and delay is not evident from (6). Refer to Figure 2 for the overall signal processing structure. Note the filter mode independent delay and scaling present in (10) or (11) are extracted from the filter modes and represented by the delay τ and weight a of the virtual source signal $s(t)$.

So far the FOS and SOS have been formulated in the Laplace domain. For implementation of these by a digital recursive filter a suitable transformation of the coefficients has to be performed. Two frequently applied methods for this purpose are the bilinear transformation and the impulse invariance method. However, in our case the corrected impulse invariance method (CIIM) [7] has to be applied since discontinuities at $t = 0$ may be present. In [6] it is shown that the digital filter coefficients of the FOS and SOS can be derived in closed form from the zeros/poles in the Laplace domain using the CIIM.

Alternatively a bilinear transformation can be used. This transformation can be performed efficiently by formulating it in terms of a 2×2 or 3×3 matrix multiplication. Hence for both the CIIM and bilinear transformation the digital filter coefficients can be computed from the zeros/poles of the FOS and SOS in the Laplace domain.

It is also evident from (10) and (11) that the zeros/poles in the Laplace domain can be computed by scaling the roots of the normalized polynomial given by $\beta_n(k)$. Hence, the coefficients of the filter modes can be computed very efficiently by pre-calculating the roots of the normalized polynomial and sorting them into FOS and SOS. The pre-calculated roots are then scaled accordingly to the desired virtual source and its parameters. After scaling the digital filter coefficients are computed by applying a CIIM or bilinear transformation. Noting that $H_0(s)$ is constant besides a delay/weight, a total of M recursive filters with ascending order from 1 to M is resulting. This solution is highly efficient compared to a realization of the filter modes by the frequently used frequency sampling method.

3.4. Practical aspects

The calculation of the roots of the normalized polynomial is prone to numerical inaccuracies for high orders n. Acceptable results have been achieved up to order $n = 75$ using MATLAB with double precision. Hence up to $P = 151$ loudspeakers can be handled straightforwardly. For higher orders advanced root finding algorithms have to be applied. Good results have been achieved in practice when storing the roots in double precision and changing

the precision to float after scaling.
For model-based rendering it is desirable to allow for time-variant source parameters, like e. g. required for the synthesis of moving virtual point sources. The proposed highly efficient computation of the filter coefficients allows to calculate these at dense time intervals. However, exchanging the filter coefficients of a recursive filter for a given state may lead to discontinuities and/or transients [8]. Care has to be taken in a real-time implementation to handle time-variant scenarios. One potential solution is to use two filters per mode m in parallel, one with the previous coefficients and one with the current ones and to cross fade between both. Due to the rapidly decaying impulse responses of the filter modes for typical scenarios the crossfade time can be chosen short.
The synthesis of multiple virtual sources can be considered by adding up the contributions from the multiple sources before the IFFT, requiring only one IFFT for multiple sources.

4. RESULTS

4.1. Implementation

The proposed approach has been implemented in MATLAB for evaluation. However, care has taken to stay in line with a potential real-time implementation. In a first step, the roots of the normalized polynomials given by $\beta_n(k)$ up to order 75 were computed using double precision, sorted into FOS/SOS and stored in a file. This file was then used to scale the roots according to (10) or (11) in order to derive the zeros/poles of the filter modes in the Laplace domain. A CIIM or bilinear transformation was then applied in order to derive the coefficients of the digital recursive filter. The FOS/SOS were realized in direct form II using single precision arithmetic. The loudspeaker signals were computed as depicted in Fig. 2. The sampling rate was $f_s = 44100$ Hz, the radius of the loudspeaker array $R = 1.5$ m, the maximum order $M = 28$ and the number of loudspeakers $P = 57$.

4.2. Evaluation results

For evaluation the frequency and phase response of the filter modes has been compared to the desired responses as given analytically by (5) and (6), respectively. The CIIM showed some deviations from the desired responses for higher orders m. The bilinear transformation provided much better results in this respect. Here the deviation from the desired magnitude frequency response was below 1 dB even for $m = 28$. Consequently for the following results the bilinear transformation has been chosen.
In a second step the overall impulse responses from the virtual source to the loudspeaker driving signals has been compared to an implementation using the frequency sampling method. Again the results showed high accuracy for the considered scenario. Results for the first two steps of evaluation are not shown due to space constraints.
Finally the synthesized sound field has been computed for the driving signals evaluated above. Figure 3 shows the resulting sound field for the synthesis of a virtual point source at position $\mathbf{x}_{\text{ps}} = [0\ 3]^T$ m. The synthesized sound field shows all typical features of NFC-HOA when comparing with the results derived in [9]. For instance, the multiple wave fronts are a result of the spatial bandlimitation and sampling. Similar results have been archived for the synthesis of a plane wave.
The evaluation of the presented approach proves its accuracy and stability in practical scenarios.

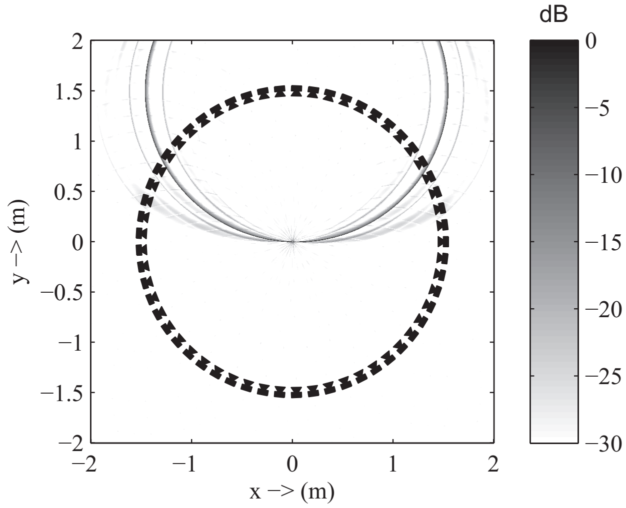

Figure 3: Spatial impulse response for the synthesis of a virtual point source at $\mathbf{x}_{\text{ps}} = [0\ 3]^T$ m using $P = 57$ loudspeakers, $M = 28$ and $f_s = 44.1$ kHz.

5. CONCLUSIONS

This paper presents an efficient parametric realization of the signal processing required for 2.5-dimensional NFC-HOA. The main building blocks are the filter modes, realized a FOS/SOS direct form II recursive filters and an IFFT. It was shown that the filter coefficients of the filter modes can be computed very efficiently from the roots of a normalized polynomial. These roots can be pre-computed and stored offline. The presented approach allows for the synthesis of time-varying acoustic scenes due to the efficient computation of the filter coefficients.
As next step the proposed technique is implemented into the SoundScape Renderer (SSR) [2], a versatile framework for model-based real-time rendering of acoustic scenes. Furthermore, extensions towards the artifact free handling of filter coefficient changes are developed.

6. REFERENCES

[1] J. Daniel, "Spatial sound encoding including near field effect: Introducing distance coding filters and a viable, new ambisonic format," in *AES 23rd International Conference*. Copenhagen, Denmark, May 2003.

[2] M. Geier, J. Ahrens, and S. Spors, "The soundscape renderer: A unified spatial audio reproduction framework for arbitrary rendering methods," in *124th AES Convention*. Amsterdam, The Netherlands, May 2008.

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

[4] J. Ahrens and S. Spors, "An analytical approach to sound field reproduction using circular and spherical loudspeaker distributions," *Acta Acustica united with Acustica*, vol. 94, no. 6, pp. 988–999, December 2008.

[5] S. Spors and J. Ahrens, "A comparison of wave field synthesis and higher-order ambisonics with respect to physical properties and spatial sampling," in *125th AES Convention*. , October 2008.

[6] H. Pomberger, "Angular and radial directivity control for spherical loudspeaker arrays," Diploma Thesis, Institute of Electronic Music and Acoustics, University of Music and Dramatic Arts, Graz, Austria, 2008.

[7] L. Jackson, "A correction to impulse invariance," *IEEE Signal Processing Letters*, vol. 7, pp. 273–275, October 2000.

[8] V. Välimäki and T. Laakso, "Suppresion of transients in time-varying recursive filters for audio signals," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Seattle, USA, 1998.

[9] J. Ahrens, H. Wierstorf, and S. Spors, "Comparison of higher order ambisonics and wave field synthesis with respect to spatial discretization artifacts in time domain," in *AES 40th International Conference on Spatial Audio*. Tokyo, Japan, October 2010.