EFFICIENT IMPLEMENTATION OF THE SPECTRAL DIVISION METHOD FOR ARBITRARY VIRTUAL SOUND FIELDS

Jens Ahrens, Mark R. P. Thomas, Ivan Tashev

Microsoft Research, One Microsoft Way, Redmond, WA 98052, USA
{*jeahrens,markth,ivantash*}*@microsoft.com*

ABSTRACT

The Spectral Division Method is an analytic approach for sound field synthesis that determines the loudspeaker driving function in the wavenumber domain. Compact expressions for the driving function in time-frequency domain or in time domain can only be determined for a low number of special cases. Generally, the involved spatial Fourier transforms have to be evaluated numerically. We present a detailed description of the computational procedure and minimize the number of required computations by exploiting the following two aspects: 1) The interval for the spatial sampling of the virtual sound field can be selected for each time-frequency bin, whereby low time-frequency bins can be sampled more coarsely, and 2) the driving function only needs to be evaluated at the locations of the loudspeakers of a given array. The inverse spatial Fourier transform is therefore not required to be evaluated at all initial spatial sampling points but only at those locations that coincide with loudspeakers.

Index Terms— Spectral Division Method, loudspeaker array, spatial Fourier transform

1. INTRODUCTION

A number of approaches for sound field synthesis exist that can be employed with linear loudspeaker arrays. Examples for analytic approaches are Wave Field Synthesis (WFS) [1] and the Spectral Division Method (SDM) [2]. An example for a numeric approach is [3]. While numeric approaches provide considerable freedom with respect to the properties of the desired sound field to be synthesized and the loudspeaker properties that can be handled, they are computationally costly and their properties are not always as predictable as those of analytic approaches. On the other hand, WFS of simple virtual sound fields can be implemented very efficiently so that real-time rendering of a complex scene composed of dozens of virtual sound sources is straightforward. However, the efficiency of WFS comes at the price of a number of approximations in the theoretical derivation.

SDM is computationally more expensive than WFS but provides a theoretically perfect result on a specific reference line. The ability of prescribing a virtual sound field on such a reference line allows for advanced applications such as [4] that are difficult to implement based on WFS or other approximative methods. It is less obvious how SDM can be implemented efficiently than it is for WFS. In this paper, we analyze in detail the computational procedure, determine appropriate settings for the involved parameters, and identify redundant and unnecessary computations.

2. SPATIAL FOURIER TRANSFORM

The spatial Fourier transform and its theorems are the building blocks of SDM presented in Sec. 3. This Section outlines the corresponding fundamentals.

We define the spatial Fourier transform as [5]

$$\tilde{S}(k_x, y, z, \omega) = \int_{-\infty}^{\infty} S(\mathbf{x}, \omega)\, e^{ik_x x}\, dx \tag{1}$$

exemplarily for the x-dimension. The corresponding inverse spatial Fourier transform is

$$S(\mathbf{x}, \omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \tilde{S}(k_x, y, z, \omega)\, e^{-ik_x x}\, dk_x\,. \tag{2}$$

The spatial Fourier transform can be defined accordingly for y and z-dimensions as well. For convenience, we consider a scenario that employs the spatial Fourier transform exclusively with respect to the x-dimension. The spatial Fourier transform can be applied to both the time domain sound field as well as the time-frequency domain sound field, which leads to the spatial spectra $\tilde{s}(k_x, y, z, t)$ and $\tilde{S}(k_x, y, z, \omega)$, respectively [6].

The k_x-domain is also referred to as *wavenumber domain*, *space-frequency domain*, *spatial frequency domain*, or k_x-*space* [7]. The parameter f_x, which is related to k_x via $k_x = 2\pi f_x$, may be termed *spatial frequency*. k_x may therefore be interpreted as *spatial radian frequency*. In order to emphasize which frequency we refer to we will speak of *time frequency* and *space frequency*, respectively.

Time-frequency spectra are usually plotted with respect to the time frequency f rather than with respect to the radian frequency $\omega = 2\pi f$. Additionally, only that half of the time-frequency spectrum that represents frequencies below half the sampling frequency is plotted. The other half is redundant because of the inherent symmetries of the real and imaginary parts. Such symmetries do not exist for spatial spectra $\tilde{S}(k_x, y, z, \omega)$ of time-frequency domain signals $S(\mathbf{x}, \omega)$ since $S(\mathbf{x}, \omega)$ is generally complex. Therefore, both halves of spatial spectra have to be plotted whereby 0 space frequency is usually put to the center of the corresponding axis. Plotting spatial spectra with respect to k_x or with respect to f_x is common, depending on the context.

3. PROBLEM FORMULATION

The Spectral Division Method (SDM) is an analytic method for sound field synthesis [2] and is applicable to planar and linear secondary source distributions. For the case of linear ones, it allows for prescribing the synthesized sound pressure along a reference
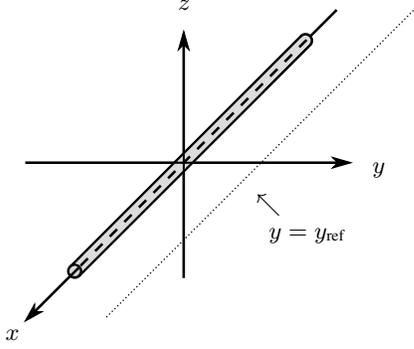
Figure 1: Illustration of the setup of a linear secondary source situated along the x-axis. The secondary source distribution is indicated by the grey shading and has infinite extent. The thin dotted line indicates the reference line (see text).

line that is parallel to the secondary source distribution and provides a perfect solution for the case of a continuous distribution of secondary sources of infinite extent. A detailed treatment of the theoretic possibilities in this context is beyond the scope of this paper. We therefore outline the theory only briefly and then concentrate on its implementation. We assume that the secondary source distribution is located along the x-axis and that the reference line is located in the horizontal plane at distance y_{ref}, i.e., $\mathbf{x}_{\text{ref}} = [x\,y_{\text{ref}}\,0]^T$. We also assume free-field conditions. Refer to Fig. 1 for an illustration of the setup.

The sound pressure $S(\mathbf{x},\omega)$ evoked by such a continuous secondary source distribution is given by an integration over the product of the driving function $D(\mathbf{x}_0,\omega)$ of the secondary source that is located at $\mathbf{x}_0 = [x_0\,0\,0]^T$ and its spatio-temporal transfer function $G(\mathbf{x},\mathbf{x}_0,\omega)$. The integration is performed along the entire secondary source contour as

$$S(\mathbf{x},\omega) = \int_{-\infty}^{\infty} D(\mathbf{x}_0,\omega)G(\mathbf{x}-\mathbf{x}_0,\omega)\,\mathrm{d}x_0 , \qquad (3)$$

Eq. (3) can be interpreted as a convolution along the x-axis [8] and the convolution theorem

$$\tilde{S}(k_x,y,z,\omega) = \tilde{D}(k_x,\omega)\cdot\tilde{G}(k_x,y,z,\mathbf{x}_0=[0\,0\,0]^T,\omega) \quad (4)$$

holds, where $\tilde{G}(k_x,y,z,\mathbf{x}_0=[0\,0\,0]^T,\omega)$ denotes the spatio-temporal transfer function of the secondary source located in the coordinate origin. The driving function $\tilde{D}(k_x,\omega)$ for the synthesis of a desired sound field $\tilde{S}(k_x,y=y_{\text{ref}},z=0,\omega)$ on the reference line $\mathbf{x}_{\text{ref}}$ can then be determined in k_x-space as [2]

$$\tilde{D}(k_x,\omega) = \frac{\tilde{S}(k_x,y=y_{\text{ref}},z=0,\omega)}{\tilde{G}(k_x,y=y_{\text{ref}},z=0,\mathbf{x}_0=[0\,0\,0]^T,\omega)} , \qquad (5)$$

which can then be transferred to time-frequency domain via (2). $\tilde{G}(\cdot)$ in (5) may not exhibit zeros, which is fulfilled, for example, for omnidirectional secondary sources.

The driving function $D(x,\omega)$ in time-frequency domain can be derived analytically for special cases of simple scenarios [9]. The only example available in the literature so far is the case of a virtual plane wave synthesized by a distribution of omni-directional secondary sources. In this case, $\tilde{S}(\cdot)$ and $\tilde{G}(\cdot)$ in (5) as well as the
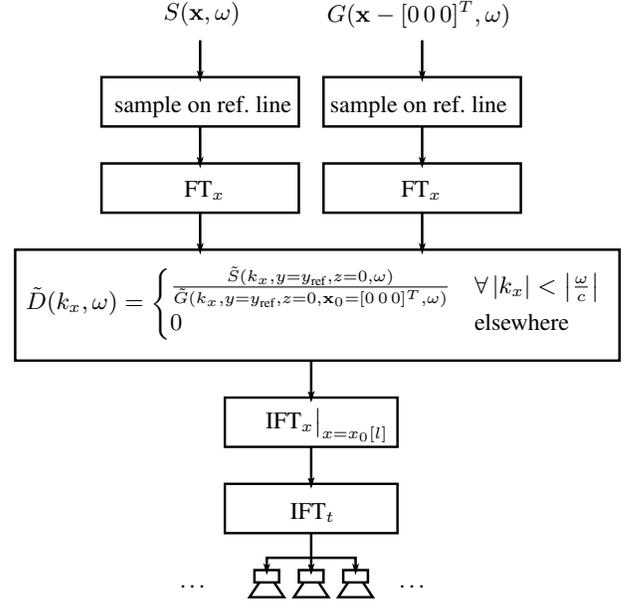

Figure 2: Block diagram of SDM implementation; FT_i: Fourier transform with respect to the indexed dimension; $x_0[l]$: x-coordinate of the l-th loudspeaker; IFT_i: inverse Fourier transform with respect to the indexed dimension

inverse spatial Fourier transform (2) of (5) can be determined analytically. The secondary source driving function $D(x,\omega)$ in time-frequency domain is then given by [2]

$$D(x,\omega) = \frac{4i\cdot e^{-ik_{\text{pw},y}y_{\text{ref}}}}{H_0^{(2)}(k_{\text{pw},y}y_{\text{ref}})}\cdot e^{-ik_{\text{pw},x}x} . \qquad (6)$$

$H_0^{(2)}(\cdot)$ denotes the Hankel function of second kind [7]. $k_{\text{pw},x}$ and $k_{\text{pw},y}$ are the components of the plane wave's wave vector in x and y-direction, respectively. The SDM solution for the considered scenario can therefore be implemented directly by designing a digital filter whose transfer function is given by (6).

Such an analytic derivation of $D(x,\omega)$ is not possible in the general case of arbitrary source-free virtual sound fields and loudspeaker radiation properties and the involved transformations have to be performed numerically. The block diagram in Fig. 2 summarizes the computational procedure. Its components will be analyzed and discussed in detail in Sec. 4 and 5.

4. FORWARD SPATIAL FOURIER TRANSFORM FOR ARBITRARY VIRTUAL SOUND FIELDS

A few selected cases of $S(\mathbf{x},\omega)$ and $G(\mathbf{x},\mathbf{x}_0=[0\,0\,0]^T,\omega)$ allow for their spatial spectra $\tilde{S}(\cdot)$ and $\tilde{G}(\cdot)$ in (5) to be determined analytically. The examples available in the literature are $S(\mathbf{x},\omega)$ being either a spherical or a plane wave and $G(\mathbf{x},\mathbf{x}_0=[0\,0\,0]^T,\omega)$ representing a spherical wave originating from the coordinate origin. In all other cases $\tilde{S}(\cdot)$ and $\tilde{G}(\cdot)$ have to be determined by numerically evaluating (1), i.e., by performing a numeric Fourier transform (also known as *discrete Fourier transform* (DFT) [8]) along the reference line $\mathbf{x}_{\text{ref}}$ for each time-frequency bin individually. The *fast*

Fourier transform (FFT) is a particularly efficient implementation of the DFT and a number of highly optimized function libraries such as [10] are available.

Note that the required number of time-frequency bins that is necessary to represent the driving function with given accuracy depends on the parameters of the considered scenario such as the length of the considered array. A general rule cannot be established.

In this Section, we discuss what parameters need to be considered in the discretization of the integral in (1) and outline how the number of computations can be minimized. We use the case of $G(\cdot)$ representing a spherical wave originating from the coordinate origin as an example. The analytical solution for $\tilde{G}(\cdot)$ in this case is [2]

$$\tilde{G}(k_x, y, z, \omega) =$$
$$\begin{cases} -\frac{i}{4} H_0^{(2)} \left(\sqrt{\left(\frac{\omega}{c}\right)^2 - k_x{}^2} \sqrt{y^2 + z^2} \right) & \text{for } |k_x| < \left|\frac{\omega}{c}\right| \\ \frac{1}{2\pi} K_0 \left(\sqrt{k_x{}^2 - \left(\frac{\omega}{c}\right)^2} \sqrt{y^2 + z^2} \right) & \text{for } \left|\frac{\omega}{c}\right| < |k_x| \ . \end{cases} \quad (7)$$

$K_0(\cdot)$ denotes the zero-th order modified Bessel function of second kind [7]. Eq. (7) is illustrated in Fig. 4(a). Note that Fig. 4(a) is symmetric with respect to $k_x = 0$ as it describes a sound field that is symmetric with respect to the y-z-plane. Such symmetry is not apparent for non-symmetric fields.

A numeric evaluation of (1) requires discretizing the integral to be a finite summation. The – preferably constant – interval Δx at which the integrand is sampled has to be chosen based on the integrand's rate of variation, i.e., the Nyquist criterion has to be satisfied so that no spatial aliasing occurs. For propagating sound fields, the Nyquist criterion requires just more than two sampling points per wavelength λ of the highest considered time-frequency [8], i.e.,

$$\Delta x < \lambda_{\min}/2 \ . \quad (8)$$

Setting Δx smaller than indicated in (8) yields redundant information and therefore causes computational overhead.

A similar criterion does not exist for evanescent sound field components, which cannot be discretized without spatial aliasing artifacts [9]. We limit our considerations to propagating virtual fields in this paper. Evanescent sound fields are to be avoided in most scenarios. Firstly, they are conceptually undesired [11] and secondly, the presence of both propagating and evanescent components in the desired field requires very high numerical precision. Note that the synthesis of a propagating field requires triggering $\tilde{G}(\cdot)$ in the range where $|k_x| < \left|\frac{\omega}{c}\right|$ whereas the synthesis of evanescent fields requires triggering $\tilde{G}(\cdot)$ in the range where $|k_x| > \left|\frac{\omega}{c}\right|$. The energy of $\tilde{G}(\cdot)$ in these two ranges differs by several orders of magnitude as evident from Fig. 4(a). The synthesis of purely evanescent fields is feasible [12]. In order to suppress virtual evanescent components we set $\tilde{D}(k_x, \omega) = 0 \forall |k_x| > \left|\frac{\omega}{c}\right|$ in (5) [9].

We need to spatially sample $S(\cdot)$ and $G(\cdot)$ at identical points along the reference line in order to be able to compute the ratio in (5). To minimize computations, we can spatially sample $S(\cdot)$ and $G(\cdot)$ for each time-frequency bin differently, whereby we set the sampling interval Δx just small enough to avoid spatial aliasing.

It is important to note that the driving function $D(x, \omega)$ has to be evaluated at the locations of the loudspeakers of a given array to determine the loudspeakers' driving signals. In order to avoid having to interpolate $D(x, \omega)$, the x-values of the sampling points should be selected such that the x-values of all loudspeaker locations are covered. Refer to Fig. 3 for an illustration.
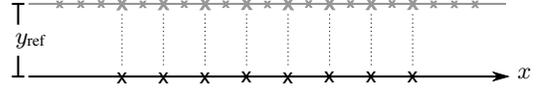

Figure 3: Relation of loudspeaker locations and spatial sampling points; black line: x-axis; gray line: reference line; black marks: loudspeaker positions; large gray marks: 'mandatory' spatial sampling points (i.e., sampling points whose x-values coincide with x-values of loudspeaker locations); small gray marks: spatial sampling points that might be required at higher frequencies or to have a wide enough window.

Similar to the sampling frequency in time-domain discretization, the spatial sampling frequency $k_{x,\mathrm{s}}$ can be established as [8]

$$k_{x,\mathrm{s}} = \frac{2\pi}{\Delta x} \ . \quad (9)$$

$k_{x,\mathrm{s}}$ is measured in the unit rad/m as is k_x. Fig. 4(b) shows a numerical result for $\tilde{G}(\cdot)$ when the sampling interval $\Delta x = 0.1$ m is too coarse to avoid aliasing at higher frequencies. In the given case, $k_{x,\mathrm{s}}$ and therefore the period of the spectral repetitions is approx. 63 rad/m, which causes spatial aliasing in the baseband $-k_{x,\mathrm{s}}/2 < k_x < k_{x,\mathrm{s}}/2$ at time frequencies higher than 1715 Hz. We assume throughout this paper that all parameters are set such that no spatial aliasing occurs.

A numeric evaluation of (1) also requires establishing finite integration boundaries, which is equivalent to windowing of the considered sound field with respect to x. For convenience, we use a rectangular window in this paper. Other types of windows are also useful [13].

Usually, it will be such that the considered sound field's amplitude decreases with distance. The wider the analysis window is the larger is the fraction of the overall energy of the sound field that is covered by the window. This does not hold for plane waves, however. Their spatial Fourier transform can be determined analytically to avoid compromises [9].

Windowing of a sound field has two major implications:

1. The energy of the sound fields spatial spectrum is smeared along the k_x-axis [13].

2. Any sound field exhibits infinite spatial extent under free-field conditions. The part of the sound field that is beyond the support of the window is excluded from the analysis.

Obviously, the longer this analysis window is the more accurate will be the result. Fig. 4(c) and 4(d) depict $\tilde{G}(\cdot)$ for different lengths of the analysis window. As expected, the shorter window causes a more significant smear of the energy. The smear is most obvious along the boundary between propagating and evanescent field components. Note that the slope around $|k_x = \omega/c|$ is significantly steeper for the longer window in Fig. 4(c) compared to the short window in Fig. 4(d). A longer window contains more sampling points and therefore requires more computations. The required accuracy and therefore the required window length will depend on the considered scenario. No general guidelines can be given.

In some scenarios it might be desired to combine analytic and numeric data. An example is the synthesis of a virtual plane or a virtual spherical wave – whose spatial Fourier transform can be determined analytically – by an array of loudspeakers with complex radiation properties [14]. In this case, $\tilde{G}(\cdot)$ has to be obtained from measurements. In order to be able to combine the analytic and the
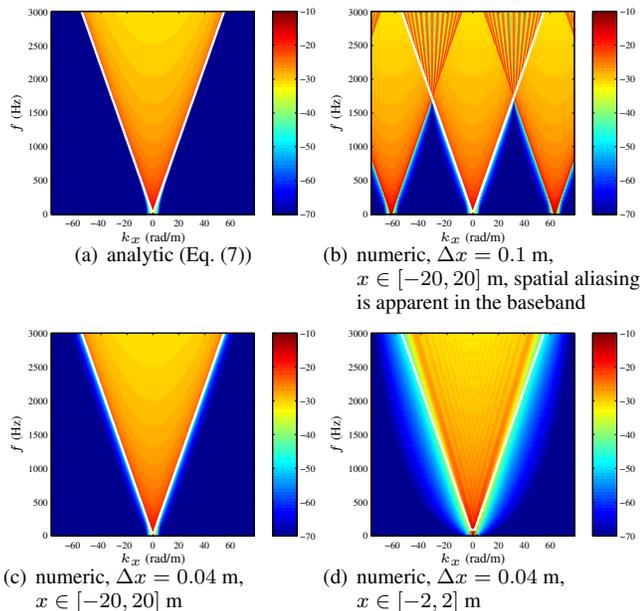
(a) analytic (Eq. (7))

(b) numeric, $\Delta x = 0.1$ m, $x \in [-20, 20]$ m, spatial aliasing is apparent in the baseband

(c) numeric, $\Delta x = 0.04$ m, $x \in [-20, 20]$ m

(d) numeric, $\Delta x = 0.04$ m, $x \in [-2, 2]$ m

Figure 4: $20 \log_{10} \left| \tilde{G}(\cdot) \right|$ for $y = 1$ m, $z = 0$; The white lines indicate $k_x = \omega/c$, i.e., the boundary between propagating and evanescent components. All values are clipped as indicated by the colorbars.

numeric data it is required that they are both available for the same values of k_x and ω. In other words, the two sets of data have to represent the same spatial and temporal sampling locations. $G(\cdot)$ should therefore be measured either continuously [15] or at appropriate discrete locations along the intended reference line. Typical implementations of the DFT compute the corresponding unique discrete values

$$k_x[l] = \frac{l}{L} k_{x,\mathrm{s}}$$
$$\forall l = -\frac{L-1}{2}, -\frac{L-1}{2} + 1, \ldots, \frac{L-1}{2} - 1, \frac{L-1}{2}, \quad (10)$$

exemplarily of uneven L and assuming that $G(\cdot)$ was measured a set of L equidistant locations with spacing Δx [8, 10]. The analytic expression for $\tilde{S}(\cdot)$ can then be evaluated at the according values of $k_x[l]$ as given by (10).

5. INVERSE SPATIAL FOURIER TRANSFORM FOR ARBITRARY VIRTUAL SOUND FIELDS

Once $\tilde{D}(k_x, \omega)$ is determined, an inverse spatial Fourier transform (2) has to be applied to yield $D(x, \omega)$. The latter is then evaluated at the positions of the loudspeakers of the array under consideration. This means that although typical implementations of the DFT give a result for all initial spatial sampling points of the time-frequency bin under consideration, we only need to compute $D(x, \omega)$ for those sampling points that coincide with loudspeaker positions (recall Fig. 3). A custom implementation of the DFT or FFT, respectively, can therefore avoid unnecessary computations.

The energy support of the time-domain representation $d(x, t)$ of the driving function will usually be preceded by silence of given duration. This silence is due to the virtual propagation of the virtual

sound field in the virtual space and therefore occurs independent of the applied method [3, 9]. The silence in $d(x, t)$ can be removed and replaced by a pure delay to reduce the resulting filter length.

6. CONCLUSIONS

We discussed the computation procedure for determining the loudspeaker driving function for the Spectral Division Method for arbitrary virtual sound fields. The involved spatial Fourier transforms have to be performed numerically. We presented criteria for selecting spatial sampling points of the virtual sound field that avoid spatial aliasing and interpolation of the resulting driving function. We showed that the virtual sound field can be sampled more coarsely at lower time frequencies than at higher ones.

We also showed that the final inverse spatial Fourier transform does not need to be evaluated for all spatial sampling points but only for those that represent loudspeaker positions. Unnecessary computations are thereby avoided.

7. REFERENCES

[1] A. J. Berkhout, D. de Vries, and P. Vogel, "Acoustic control by wave field synthesis," *JASA*, vol. 93, no. 5, pp. 2764–2778, May 1993.

[2] J. Ahrens and S. Spors, "Sound field reproduction using planar and linear arrays of loudspeakers," *IEEE Trans. on Audio, Sp., and Lang. Proc.*, vol. 18, no. 8, pp. 2038 – 2050, Nov. 2010.

[3] M. Kolundžija, C. Faller, and M. Vetterli, "Reproducing Sound Fields Using MIMO Acoustic Channel Inversion," *JAES*, vol. 59, no. 10, pp. 721–734, Oct. 2011.

[4] J. Ahrens, M. R. P. Thomas, and I. Tashev, "Gentle Acoustic Crosstalk Cancelation using the Spectral Division Method and Ambiophonics," in *IEEE WASPAA*, New Paltz, NY, Oct. 2013.

[5] R. N. Bracewell, *The Fourier Transform and Its Applications*. Singapore: McGraw-Hill, 2000.

[6] R. Rabenstein, P. Steffen, and S. Spors, "Representation of two-dimensional wave fields by multidimensional signals," *Elsevier App. Sp. and Audio Proc.*, vol. 86, no. 6, pp. 1341 – 1351, June 2006.

[7] E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustic Holography*. London: Academic Press, 1999.

[8] B. Girod, R. Rabenstein, and A. Stenger, *Signals and Systems*. New York: Wiley, 2001.

[9] J. Ahrens, *Analytic Methods of Sound Field Synthesis*. Berlin/Heidelberg: Springer, 2012.

[10] M. Frigo and S. G. Johnson, "FFTW," http://www.fftw.org, accessed May 9, 2013.

[11] S. Spors and J. Ahrens, "Reproduction of Focused Sources by the Spectral Division Method," in *Proc. IEEE ISCCSP*, Limassol, Cyprus, 2010, pp. 1 – 5.

[12] H. Itou, K. Furuya, and Y. Haneda, "Evanescent wave reproduction using linear array of loudspeakers," in *Proc. IEEE WASPAA*, New Paltz, NY, USA, 2011, pp. 37 – 40.

[13] F. J. Harris, "On the use of windows for harmonic analysis with the discrete fourier transform," *Proceedings of the IEEE*, vol. 66, no. 1, pp. 51 – 83, Jan. 1978.

[14] J. Ahrens and S. Spors, "An analytical approach to 2.5D sound field reproduction employing linear distributions of non-omnidirectional loudspeakers," in *IEEE ICASSP*, Dallas, Texas, Mar. 2010, pp. 105–108.

[15] G. Enzner, "3D-continuous-azimuth acquisition of head-related impulse responses using multi-channel adaptive filtering," in *IEEE WASPAA*, New Paltz, NY, USA, Oct. 2009, pp. 325–328.