Audio Engineering Society

Convention e-Brief 657

Presented at the 151st Convention
2021 October, Online

Real-time Implementation of the Spectral Division Method for Binaural Personal Audio Delivery with Head Tracking

Yue Qiao[1] and Edgar Choueiri[1]

[1] *3D Audio and Applied Acoustics Laboratory, Princeton University, Princeton, New Jersey, 08544, USA*

Correspondence should be addressed to Yue Qiao (yqiao@princeton.edu)

ABSTRACT

A scheme for implementing the Spectral Division Method (SDM) in real time for delivering binaural personal audio to multiple listeners with head tracking is presented. The SDM, as an analytical approach for sound field reproduction, has been applied to generating personal audio filters that create acoustically bright and dark zones. However, only the case of static listening positions has been investigated. In realistic situations, the performance of such personal audio delivery systems will degrade significantly when the listeners move out of the "sweet spots". In order to achieve dynamic personal audio delivery that compensates for listeners' head movements, the SDM-based filters are updated in real time through simple multiplications in the wavenumber domain, by utilizing the shifting theorem of the spatial Fourier transform along the x-axis. Furthermore, by selecting two spatial window functions targeted at two ears, the generated filters are able to deliver separate binaural personal audio to multiple listeners. The proposed scheme offers an intuitive and efficient solution for binaural personal audio delivery with head tracking, without the need of regenerating multiple filters as required by traditional methods.

1 Introduction

The delivery of personal audio to multiple listeners within the same physical space using loudspeakers has received much research attention since it was first proposed in the 1990s [1]. In order to achieve acoustical separation between listeners, personal audio filters are usually generated by using either the Pressure Matching (PM) method [2], where the difference between the specified target sound pressures at the ears and the actual sound pressures generated by the loudspeakers is minimized, or the Acoustic Contrast Control (ACC) method [3], where the difference between sound energy in acoustically "bright" and "dark" listening zones is maximized. Both methods would require additional regularization parameters to be specified in problem formulation [4] to limit loudspeaker gains and ensure the robustness of filters, which can degrade the performance if not chosen properly.

As an alternative approach that does not require regularization, Okamoto [5] employed the Spectral Division Method (SDM) for generating bright and dark zones and validated its effectiveness in [6]. The SDM is an analytical approach for sound field synthesis proposed by Ahrens and Spors [7] based on the spatial Fourier transform. For personal audio problems, given the target distribution of sound pressures in the bright and dark zones, SDM generates the loudspeaker driving function for accurately reproducing the specified target. In [5], SDM is applied for generating a *single* bright zone with a pre-specified spatial window as the target pressure distribution. We show in this e-brief that by using *two* spatial windows to separately "illuminate" the two ears of a listener, the binaural personal audio delivery can be achieved with different audio programs assigned to each ear.

For personal audio delivery designed for static listening positions, the audio separation level can drop significantly when listeners move out of the "sweet spots". Therefore, it is highly desirable to implement dynamic reproduction by updating filters corresponding to the listeners' head positions in real time. For approaches based on PM, Gálvez et al. [8] proposed a method for dynamic reproduction by deriving the analytical form of the PM filter coefficients and adjusting their gain and delay components in real time. For the SDM-based approach [5], however, only the case of static sound zone configuration is implemented. In this e-brief, we present a scheme for dynamically updating the SDM-based filters as a function of head movements in real time, which enables binaural personal audio delivery to multiple listeners with head tracking.

The rest of this e-brief is organized as follows: Section 2 briefly introduces the relevant aspects of SDM theory and the spatial shifting operation, which is essential to the implementation of the proposed scheme; Section 3 details the actual real-time implementation of the proposed scheme; Section 4 discusses the main benefits of the proposed scheme; Section 5 summarizes the main findings of this e-brief, as well as potential future improvements.

2 SDM and Target Shifting

In sound field reproduction, given a continuous sound source distribution with driving function $D(\mathbf{x}_0, \omega)$, where $\mathbf{x}$ is the position vector and ω is the frequency, we can derive the sound pressure distribution at target position, $P(\mathbf{x}, \omega)$, once the spatial-temporal transfer function $G(\mathbf{x} - \mathbf{x_0}, \omega)$ is known. For a continuous linear source array distributed along the x-axis (i.e., $\mathbf{x_0} = [x_0, 0, 0]^T$), the target sound pressure at $\mathbf{x} = [x, y, 0]^T$ is given by [7]

$$P_y(\mathbf{x}, \omega) = \int_{-\infty}^{\infty} D(\mathbf{x_0}, \omega) G_y(\mathbf{x} - \mathbf{x_0}, \omega) dx_0. \quad (1)$$

By performing the spatial Fourier transform along the x-axis and applying the convolution theorem, the following relation is obtained as

$$\tilde{P}_y(k_x, y, 0, \omega) = \tilde{D}(k_x, \omega) \cdot \tilde{G}_y(k_x, y, 0, \omega), \quad (2)$$

where the tilde symbol denotes the spatial Fourier transform with respect to the x direction, and k_x is wavenumber component in the x-direction. Using the equation above, once the target distribution $P(\mathbf{x}, \omega)$ is specified, SDM is applied to derive the source driving function by division in the wavenumber domain at each frequency ω, which results in filter coefficients corresponding to each source position.

In the context of personal audio delivery, when listeners move out of the "sweet spots", the filter coefficients need to be updated to compensate for head movements. Assuming the listener moves from $[0, y, 0]^T$ to $[x_0, y, 0]^T$ (corresponding to the left/right head movement), the specified target can be shifted accordingly as

$$P_{shifted}(x, \omega) = P(x - x_0, \omega), \quad (3)$$

whereas in the wavenumber domain, the shifting is represented as

$$\tilde{P}_{shifted}(k_x, \omega) = e^{-jk_x x_0} \tilde{P}(k_x, \omega). \quad (4)$$

The equivalence above is crucial to the proposed scheme. Without having to regenerate filters at each new listener position, it is possible to only compute filters for the center position at first, and then update the coefficients by multiplying complex exponential in the wavenumber domain and transforming back to the time-frequency domain using the inverse Fourier transform. This leads to significant simplification of the dynamic filter updating process.

3 Practical Implementation

In actual implementation of SDM for personal audio delivery, the first step is to specify the target distribution as a spatial window centered at the listener's head position. In [6], a single Rectangular/Hanning spatial window is used for the bright zone. Here, as shown in Figure 1, we specify two Hanning windows separately centered at each ear of the target listener, with the goal of creating both acoustic contrast between listeners and crosstalk cancellation for a single listener. A similar setup using two Rectangular windows to provide crosstalk cancellation only is seen in [9]. The

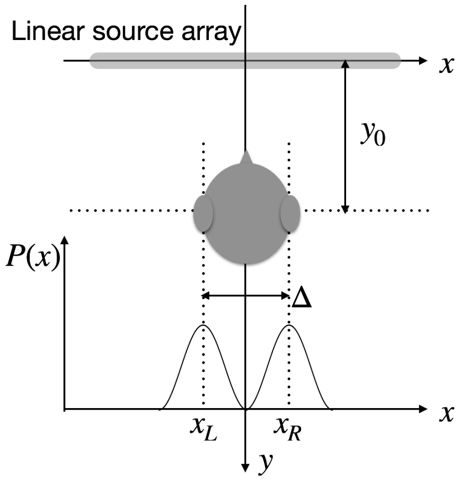

Fig. 1: Illustration of the specified target pressure distribution. Two spatial Hanning windows are centered at the positions of listener's ears (x_L, x_R). The listener is located at $[0, y_0, 0]^T$ as reference.

target pressure distribution $P(x,\omega)$ with two Hanning windows is given by

$$P(x,\Delta) = P_{Hann}(x - x_L, \Delta) + P_{Hann}(x - x_R, \Delta), \quad (5)$$

where Δ is the width of the spatial window (also the head width), and x_L and x_R denote the positions of the left and right ears. The spatial Hanning window function $P_{Hann}(x,\Delta)$ is defined as

$$P_{Hann}(x,\Delta) = \begin{cases} \cos^2(\frac{\pi x}{\Delta}), & |x| \leq \Delta/2, \\ 0, & |x| > \Delta/2. \end{cases} \quad (6)$$

In practice where the continuous linear source is replaced by an array of finite, discrete loudspeakers, the analytically-derived solution needs to be adapted to discrete form and modified to generate stable audio filters. First, as the solution from Equation 2 contains evanescent components for $|k_x| > \omega/c$ that are negligible at moderate distances [7], these terms are set to zero during actual computation. Second, in order to apply filters to a linear array of loudspeakers, the derived source driving function is subject to truncation and discretization, the effects of which have been discussed in [7].

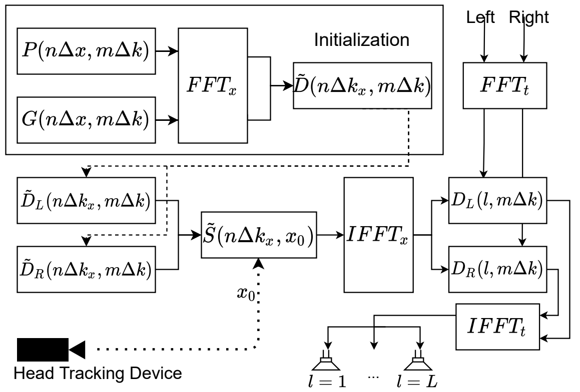

Fig. 2: Block diagram for the implementation. All quantities are represented in discrete form. The truncation and discretization processes are neglected.

The block diagram for implementing the proposed scheme is shown in Figure 2. Two separate sets of FFTs/IFFTs are applied in the time and space domains, each with length N_t and N_x. For FFT computation in the space domain, a spatial grid along the x-axis is defined as $x = n\Delta x$, where $n = -N/2, \cdots, N/2-1$; the choice of Δx should follow $\Delta x < \lambda_{min}/2$ to avoid spatial aliasing, where λ_{min} is the minimum wavelength of interest.

For initialization, the target distribution P and the spatial-temporal transfer function G are transformed into the space domain to obtain the the source driving function $\tilde{D}$. Then, the source driving function for centered listener position is shifted towards each of the two ears to produce $\tilde{D_L}$ and $\tilde{D_R}$ by multiplying complex exponential terms $e^{jk_x\Delta/2}$ and $e^{-jk_x\Delta/2}$. During the audio playback, a shifting vector $\tilde{S} = e^{-jk_x x_0}$ is updated with head position x_0 received from the head tracking device in real time, and then multiplied with $\tilde{D_L}$ and $\tilde{D_R}$ to produce new coefficients. The discrete source driving function is then obtained through $IFFT_x$. Finally, the filter coefficients corresponding to the actual position of L loudspeakers are selected and convolved with the left and right input signals in the frequency domain to generate binaural personal audio for the target listener.

4 Discussion

The proposed scheme has many advantages for delivering binaural personal audio. First, the generated SDM filters provide fixed sound images to the listener irrespective of head movements, making it possible to cascade other types of audio filters designed only for the static head position. For example, a static crosstalk cancellation filter for a single listener could be added to further increase the crosstalk cancellation level at all positions, and a static equalization filter could be applied to correct the introduced spectral coloration. Second, as is pointed out in [10], in generating the SDM filters, it is possible to use measured spatial-temporal transfer functions $\tilde{G}$ instead of analytical ones. This refines the acoustical model and is expected to improve the acoustic contrast. In addition, the resulting target sound field is entirely controlled by the specified spatial window, offering great flexibility for different scenarios, such as a single, wider window for more robustness, or two windows with varying width/position for different listeners.

5 Conclusion

In this e-brief, we propose a scheme for implementing the Spectral Division Method in real time for delivering binaural personal audio to multiple listeners with head tracking. The filter coefficients are updated in the wavenumber domain by multiplying the complex exponential terms corresponding to the tracked head positions, and transformed back to the space domain. With the specified two Hanning spatial windows, a fixed binaural sound image is created for the listener irrespective of head positions, which expands the sweet spot and increases the robustness of the personal audio system.

In future work, the proposed scheme for updating filters could be further optimized for DSP performance. In particular, Ahrens et al. [10] suggested a few processes to improve the efficiency of SDM filter calculation by choosing a frequency-dependent spatial sampling and performing the inverse spatial Fourier transform only for the loudspeaker positions; both of these operations could be implemented. Furthermore, numerical simulation and physical experiments with loudspeaker arrays would be essential to evaluating the performance of the proposed scheme, and comparing to that of other existing approaches.

References

[1] Druyvesteyn, W. F. and Garas, J., "Personal Sound," *Journal of the Audio Engineering Society*, 45(9), pp. 685–701, 1997.

[2] Kirkeby, O., Nelson, P., Hamada, H., and Orduna-Bustamante, F., "Fast Deconvolution of Multichannel Systems Using Regularization," *IEEE Transactions on Speech and Audio Processing*, 6(2), pp. 189–194, 1998, ISSN 1558-2353, doi: 10.1109/89.661479.

[3] Choi, J.-W. and Kim, Y.-H., "Generation of an Acoustically Bright Zone with an Illuminated Region Using Multiple Sources," *The Journal of the Acoustical Society of America*, 111(4), pp. 1695–1700, 2002, ISSN 0001-4966, doi: 10.1121/1.1456926.

[4] Coleman, P., Jackson, P. J. B., Olik, M., Møller, M., Olsen, M., and Abildgaard Pedersen, J., "Acoustic Contrast, Planarity and Robustness of Sound Zone Methods Using a Circular Loudspeaker Array," *The Journal of the Acoustical Society of America*, 135(4), pp. 1929–1940, 2014, ISSN 0001-4966, doi:10.1121/1.4866442.

[5] Okamoto, T., "Generation of Multiple Sound Zones by Spatial Filtering in Wavenumber Domain Using a Linear Array of Loudspeakers," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4733–4737, 2014, ISSN 2379-190X, doi: 10.1109/ICASSP.2014.6854500.

[6] Okamoto, T. and Sakaguchi, A., "Experimental Validation of Spatial Fourier Transform-Based Multiple Sound Zone Generation with a Linear Loudspeaker Array," *The Journal of the Acoustical Society of America*, 141(3), pp. 1769–1780, 2017, ISSN 0001-4966, doi:10.1121/1.4977995.

[7] Ahrens, J. and Spors, S., "Sound Field Reproduction Using Planar and Linear Arrays of Loudspeakers," *IEEE Transactions on Audio, Speech, and Language Processing*, 18(8), pp. 2038–2050, 2010, ISSN 1558-7924, doi:10.1109/TASL.2010.2041106.

[8] Gálvez, M. F. S., Menzies, D., and Fazi, F. M., "Dynamic Audio Reproduction with Linear Loudspeaker Arrays," *Journal of the Audio Engineering Society*, 67(4), pp. 190–200, 2019, ISSN 15494950, doi:10.17743/jaes.2019.0007.

[9] Ahrens, J., Thomas, M. R. P., and Tashev, I., "Gentle Acoustic Crosstalk Cancelation Using

the Spectral Division Method and Ambiophonics," in *2013 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 1–4, 2013, ISSN 1947-1629, doi:10.1109/WASPAA.2013.6701827.

[10] Ahrens, J., Thomas, M. R. P., and Tashev, I., "Efficient Implementation of the Spectral Division Method for Arbitrary Virtual Sound Fields," in *2013 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 1–4, 2013, ISSN 1947-1629, doi:10.1109/WASPAA.2013.6701828.