LOW FREQUENCY CROSSTALK CANCELLATION AND ITS RELATIONSHIP TO AMPLITUDE PANNING

Eric Carlos Hamdan   Filippo Maria Fazi

Institute of Sound and Vibration Research, University of Southampton, SO17 1BJ, United Kingdom

ABSTRACT

The low frequency behaviour of the source strength solution of the two loudspeaker crosstalk cancellation system is analysed. Using the rigid sphere head model, an accurate 1^{st} order approximation of the source strength solution is derived, assuming a centred listener position. It is shown that the 1^{st} order approximation of the source strength solution corresponds to the well-known stereo sine law. The validity of this approximation for real systems is supported by a comparison of simulated low frequency complex source strengths and the real stereo panning functions. To obtain a realistic analysis, the simulated source strengths are created using KEMAR head-related transfer functions. It is shown that the simulated source strengths converge to the stereo sine law in the low frequency limit under the expected conditions.

Index Terms— crosstalk cancellation, stereo, binaural, amplitude panning

1. INTRODUCTION

Two loudspeaker crosstalk cancellation (CTC) systems were proposed nearly sixty years ago as a means to reproduce binaural audio at a listener's ears *without* headphones and is still an active area of research today [1, 2]. The theory underpinning the classic CTC system can be described by a linear inverse problem relating field pressures to individual source strengths transformed by a radiation matrix [3]. The inversion of acoustic transfer functions between the loudspeakers and the ears contained in the radiation matrix allows for reproduction of a desired binaural audio stream without crosstalk, in ideal conditions.

As the inverse of the radiation matrix is the key ingredient to achieve CTC, it has been an object of study in the context of CTC systems, and more generally sound field control, for several decades [4, 5, 6, 7, 8]. It has been shown by some authors that, depending on the assumed acoustic transfer functions and problem geometry, the radiation matrix to be inverted will be ill-conditioned for particular frequencies of reproduction [9, 10]. Ill-conditioned CTC systems suffer from sensitivity to small perturbations in the system, e.g., a deviation in assumed acoustic environment or loudspeaker misalignment [11]. This ill-conditioning can be paired with a large inverse norm. Indeed, in the low frequencies the radiation matrix inverse norm can be very large in the case of compact CTC systems [12]. Large norms in general can potentially result in loss of system dynamic range and audio quality [6].

Regularisation is often used to improve system robustness to perturbations and reduce the inverse norm so that the final loudspeaker signals are tolerable in magnitude [11]. A prevailing design pattern has been to find a compromise between the radiation matrix inverse norm and the errors introduced by regularisation [3]. Alternatively, or in combination, the geometry of the system and specification of the radiation matrix may be altered. For example, Takeuchi and Nelson proposed the Optimal Source Distribution (OSD), which suggests a continuum of stereo loudspeaker pairs of increasing span with decreasing frequency in order to achieve an inverse filter with a flat magnitude response and a system condition number of one [6].

Much of the previous research has focused on the radiation matrix inverse norm and system condition as quantities around which to design a desired CTC system. While these methods anticipate the inclusion of a desired binaural signal, they have not explicitly analysed the general source strength solution to the CTC radiation problem as stated in (5) below. It was noted briefly by Takeuchi and Nelson that the end reproduction is highly dependent on the desired binaural signal, however this idea was not explored in detail [6]. Instead, the focus has been directed towards the behaviour of the radiation matrix inverse. Therefore, the motivation driving this work is to shed further light on the source strength solution.

The following study will analyse the 1^{st} order approximation of the CTC source strength solution, i.e., low frequency approximation. It will be shown that under a given set of reasonable assumptions the 1^{st} order approximation of the source strengths amounts to ratios of scaled phase differences. In order to offer an accurate analysis in regards to the presence of a human head, the rigid sphere head model is introduced. It has been shown previously that the rigid sphere is a valid model for the human head at low frequencies in anechoic conditions [13]. It is shown that in the low frequency limit, the analytical complex source strengths converge to the real-valued stereo sine law when the radiation matrix and desired binaural signal are commonly derived from the rigid sphere HRTF assuming plane wave sources. This specific result corresponds to the case of *in situ* measurement and reproduction of a single plane wave virtual source at low frequencies. It is then shown that source strengths numerically calculated from far field mannequin HRTF data converge to the stereo sine law in the low frequencies with a decreasing imaginary component, confirming that the stereo sine law is a valid 1st order approximation of the CTC source strengths under the stated assumptions.

2. LOW FREQUENCY RIGID SPHERE HRTF

It is well known that the 1st order approximation of the rigid sphere HRTF due to a far field source is [14]

$$H(k, a, \Theta_{L,R}) \approx 1 - j\frac{3}{2}ka\cos\Theta_{L,R}, \quad ka \ll 1, \tag{1}$$

where $k = \frac{2\pi}{\lambda}$ is the wavenumber, a is the head radius in metres and $\Theta_{L,R}$ are the angles between the incident plane wave arrival vector and the left and right ear position vectors. These vectors are measured from the origin that is placed at the centre of the listener's head. From this it can be said $|H(k, a, \Theta_{L,R})| \approx 1$ and $\angle H(k, a, \Theta_{L,R}) \approx -\frac{3}{2}ka\cos\Theta_{L,R}$ for $ka \ll 1$, i.e., low frequencies.

Note that $ka\cos\Theta_{L,R} = k\|\hat{\mathbf{n}}_S\|_2\|\mathbf{x}_{L,R}\|_2\cos\Theta_{L,R} = k\hat{\mathbf{n}}_S \cdot \mathbf{x}_{L,R} = \mathbf{k}_S \cdot \mathbf{x}_{L,R}$, where $\hat{\mathbf{n}}_S, \mathbf{x}_{L,R} \in \mathbb{R}^3$ are the plane wave direction of arrival and ear position vectors, respectively. Furthermore, the ears are assumed to be diametrically opposed. Then it can be said $\mathbf{x}_L = -\mathbf{x}_R$. This allows the left and right HRTFs to be expressed in terms of the left ear position vector only as

$$H(k, a, \Theta_L) \approx 1 - j\frac{3}{2}\mathbf{k}_S \cdot \mathbf{x}_L, \tag{2}$$

$$H(k, a, \Theta_R) \approx 1 + j\frac{3}{2}\mathbf{k}_S \cdot \mathbf{x}_L, \tag{3}$$

valid for $ka \ll 1$.

(2) and (3) provide a simple model for approximating the human HRTF at low frequencies [14].

3. CROSSTALK CANCELLATION SYSTEM

The design of the two channel CTC system can be expressed in the frequency domain as a linear system of the form [15]

$$\mathbf{p} = \mathbf{G}\mathbf{H}\mathbf{d}, \tag{4}$$

where $\mathbf{p} \in \mathbb{C}^2$ is the reproduced binaural signal at the ears (called *control points*), $\mathbf{d} \in \mathbb{C}^2$ is the desired binaural signal input, $\mathbf{G} \in \mathbb{C}^{2\times 2}$ is the matrix of acoustic transfer functions between the plane wave sources and the ears (called the *radiation matrix*) and $\mathbf{H} \in \mathbb{C}^{2\times 2}$ are the CTC inverse filters, such that $\mathbf{GH} = \mathbf{I}_2 e^{-j\omega\Delta t}$ (if no regularisation is applied). If $\mathbf{G}$ is full rank, $\mathbf{H} = \mathbf{G}^{-1}e^{-j\omega\Delta t}$. Note that the delay $e^{-j\omega\Delta t}$ is necessary for a causal solution but will omitted for brevity in the following analysis, although its presence is implied.

4. LOW FREQUENCY SOURCE STRENGTHS

The source strength solution will now be defined as

$$\mathbf{q} = \mathbf{H}\mathbf{d}. \tag{5}$$

To obtain a general and accurate low frequency approximation for $\mathbf{q}$ that is valid for the centred listening position and allows for head rotation, the general form of the radiation matrix and desired binaural signal are defined as

$$\mathbf{G} = \begin{bmatrix} G_1 & G_2 \\ G_1^* & G_2^* \end{bmatrix}, \quad \mathbf{d} = \begin{bmatrix} d \\ d^* \end{bmatrix}. \tag{6}$$

This specification matches the conjugate symmetry of the low frequency rigid sphere HRTF (Section 2) and is therefore a reasonable assumption.

The expression for $\mathbf{q}$ is readily found as

$$\mathbf{q} = \frac{1}{\Im\left(G_1 G_2^*\right)} \begin{bmatrix} \Im\left(G_2^* d\right) \\ \Im\left(G_1 d^*\right) \end{bmatrix}. \tag{7}$$

While (7) may be valid for plane waves incident on a shadowless head, it is *not* in general for a human head until the 1st order (low frequency) approximation is introduced. Considering that the arbitrary complex number $z \in \mathbb{C}$ has the 1st order approximation $z \approx |z|\left(1 + j\angle z\right), \angle z \ll 1$, the 1st order approximation of (7) is readily found as

$$\mathbf{q} \approx \alpha \begin{bmatrix} |G_2|\left(\angle d - \angle G_2\right) \\ |G_1|\left(\angle G_1 - \angle d\right) \end{bmatrix}, \quad \angle G_1, \angle G_2, \angle d \ll 1, \tag{8}$$

where $\alpha \equiv \frac{|d|}{|G_1||G_2|(\angle G_1 - \angle G_2)}$. Note that $\angle G_1, \angle G_2, \angle d \ll 1$ is satisfied by $ka \ll 1$ in the CTC problem. This result shows that $q_1, q_2 \in \mathbb{R}$ are scaled phase differences at low frequencies.

From (8) several observations can be made. In general, as the ipsilateral and contralateral source-to-control point transfer paths become increasingly close in phase, e.g., the stereo span decreases, the signals become increasingly large in magnitude until reaching a singularity when the loudspeakers are co-located or equidistant to each ear. In addition, when the phase of the desired binaural signal matches the phase of the ipsilateral or contralateral source-to-control point transfer paths, the opposite loudspeaker is off. In these cases, the remaining loudspeaker becomes a physical source and should yield the most stable virtual image, completely robust to head movements and translation (ignoring loudspeaker directivity).

Furthermore, at low frequencies q_1 and q_2 are sensitive to the ratio of magnitudes $\frac{|d|}{|G_1|}$ and $\frac{|d|}{|G_2|}$, respectively. If G_1, G_2, d are taken to be rigid sphere HRTFs, the magnitudes will ideally tend to 1 at low frequencies and $\mathbf{q}$ will be scaled only by the inverse phase difference, or path length difference, between the sources and the left or right ear.

In the following section, the particular solution for (8) considering a rigid sphere head model will be analysed and it will be shown that this solution is the stereo sine law when the HRTFs are derived from the rigid sphere.

5. LOW FREQUENCY SOURCE STRENGTHS CONSIDERING A RIGID SPHERE

In the following, the incident sound field is represented by $e^{-j\mathbf{k}\cdot\mathbf{x}}$, i.e., a plane wave, for both the target virtual source and secondary sources (loudspeakers). Using the 1st order approximation for the rigid sphere HRTF given in Section 2, the radiation matrix and desired binaural signals for $ka \ll 1$ are then specified as

$$\mathbf{G} \approx \begin{bmatrix} 1 - j\frac{3}{2}\mathbf{k}_L \cdot \mathbf{x}_L & 1 - j\frac{3}{2}\mathbf{k}_R \cdot \mathbf{x}_L \\ 1 + j\frac{3}{2}\mathbf{k}_L \cdot \mathbf{x}_L & 1 + j\frac{3}{2}\mathbf{k}_R \cdot \mathbf{x}_L \end{bmatrix}, \tag{9}$$

and

$$\mathbf{d} \approx \begin{bmatrix} 1 - j\frac{3}{2}\mathbf{k}_S \cdot \mathbf{x}_L \\ 1 + j\frac{3}{2}\mathbf{k}_S \cdot \mathbf{x}_L \end{bmatrix}, \tag{10}$$

respectively. It follows that the low frequency source strengths for $ka \ll 1$ are

$$\mathbf{q} \approx \frac{1}{\mathbf{k}_R \cdot \mathbf{x}_L - \mathbf{k}_L \cdot \mathbf{x}_L} \begin{bmatrix} \mathbf{k}_R \cdot \mathbf{x}_L - \mathbf{k}_S \cdot \mathbf{x}_L \\ \mathbf{k}_S \cdot \mathbf{x}_L - \mathbf{k}_L \cdot \mathbf{x}_L \end{bmatrix}. \tag{11}$$

Both the frequency dependent wavenumber k and the head radius term a cancel out, leaving the result

$$\mathbf{q} \approx \frac{1}{\cos\Theta_{RL} - \cos\Theta_{LL}} \begin{bmatrix} \cos\Theta_{RL} - \cos\Theta_{SL} \\ \cos\Theta_{SL} - \cos\Theta_{LL} \end{bmatrix}. \tag{12}$$

The final simplification is to consider the horizontal plane only, so that in terms of ISO spherical coordinates, (θ, ϕ), the inclination $\theta = \frac{\pi}{2}$ [16]. Note that in the horizontal plane $\cos\Theta_{LL} = \cos\left(\phi_L - \phi_L^{'}\right)$ where ϕ_L and $\phi_L^{'}$ are the left-most plane wave angle of arrival and the left ear angle, respectively, measured from the x axis [17]. This can be expressed as $\cos\Theta_{LL} = -\sin\left(\gamma_L - \phi_L^{'}\right)$, where positive γ_L is now taken to be the plane wave azimuth displacement from the listener's mid-sagittal plane in the counter-clockwise direction. Also note $\gamma_L = -\gamma_R$ for the symmetric two channel loudspeaker configuration.

With these further simplifications, for $ka \ll 1$ the following important relationship is obtained:

$$\mathbf{q} \approx \frac{1}{\sum_{n=1}^{2} g_n} \begin{bmatrix} \sin\left(\gamma_L + \phi_L^{'}\right) + \sin\left(\gamma_S - \phi_L^{'}\right) \\ \sin\left(\gamma_L - \phi_L^{'}\right) - \sin\left(\gamma_S - \phi_L^{'}\right) \end{bmatrix}, \tag{13}$$

which is in fact the stereo sine law normalised by the sum of the loudspeaker gains, where the gains are defined as $g_1 \equiv \sin\left(\gamma_L + \phi_L^{'}\right) + \sin\left(\gamma_S - \phi_L^{'}\right)$ and $g_2 \equiv \sin\left(\gamma_L - \phi_L^{'}\right) - \sin\left(\gamma_S - \phi_L^{'}\right)$ [18]. This solution conveniently accounts for head rotation which can be exploited in an adaptive system, as is done for Compensated Amplitude Panning (CAP) [19].

The result given by (13) shows that the 1st order approximation of the CTC source strengths is the stereo sine law when the radiation matrix and desired binaural signal commonly originate from the 1st order far field rigid sphere HRTFs. Based on this finding, and the established accuracy of the rigid sphere head model at low frequencies, it is expected that the low frequency source strengths derived from mannequin HRTFs in the same manner will also converge to the stereo sine law in the low frequencies.

The following section uses measured data to show that the source strengths created from mannequin HRTFs do converge to the stereo sine law for virtual source angles in the horizontal plane under the stated assumptions. Note that in this work, only the forward facing case is analysed in support of brevity.

6. EXPERIMENTAL VERIFICATION

To verify that the stereo sine law is a 1st order approximation for the source strengths created using human far field HRTFs, KEMAR HRTFs openly available from the Institut fur Technische Akustik at Technische Universitat Berlin (TU-Berlin) were used [20]. The chosen HRTFs were measured at a distance of two metres from the centre of the KEMAR head [20].

For the simulation, the two channel CTC loudspeaker geometry under consideration was chosen to match the classic stereo configuration such that the left and right loudspeakers were placed at angles $\gamma_L = 30°$ and $\gamma_R = 330°$, respectively. Accordingly, the KEMAR HRTFs corresponding to those angles were used in the formulation of $\mathbf{G}$ to approximate plane waves incident on the mannequin head from those directions. Source strengths for a range of virtual source angles $\gamma_S \in [0°, 360°)$ were calculated at one degree increments also using the corresponding HRTFs from the KEMAR measurements for $\mathbf{d}$. These source strengths, $\mathbf{q}$, were calculated according to (5).

The following analysis will focus on the low frequency behaviour of the calculated $\mathbf{q}$ signals.

7. EXPERIMENTAL RESULTS

Figure 1 shows the real and imaginary components of the calculated $\mathbf{q}$ signals for $f \approx 52$ Hz, over the range of virtual source angles $\gamma_S \in [0°, 360°)$. Overlaid is the stereo sine law (for zero head rotation) calculated for the same set of virtual source angles.

The results show that the real components tend to the stereo panning signals, L and R (see (13)), and that the imaginary components are near zero, as expected. Also, there is symmetry about the interaural axis associated with the rigid sphere model.

For comparison, Figures 2 and 3 show the behaviour of q_1 (the left loudspeaker (30°) signal) over a larger range of frequencies, $f \in [52, 1000]$ Hz. The two surface plots are used to show the magnitude (in dB) of the real and imaginary components of q_1 as functions of frequency and virtual source angle γ_S, respectively. Magnitudes below -40 dB appear black, and below -80 dB appear white. Data for q_2 was omitted for brevity, but it exhibited identical trends mirrored about the KEMAR's mid-sagittal plane.

Figure 2 shows that the real component maintains a similar envelope across frequency except for a gradual roll-off around 400 Hz for virtual sources placed around 90°. As expected from (8), the magnitude is zero for the virtual source co-located with the opposite loudspeaker (330°). For the opposite loudspeaker position mirrored about the interaural axis (210°), q_1 is also low in magnitude across frequency (this trend shifts inward in angle closer to 1000 Hz). Even up to 1000 Hz a single loudspeaker provides most of the energy when the virtual source is co-located with the loudspeaker front and rear mirror positions.

Figure 3 gives some insight into this phenomenon. When the virtual source is co-located with the physical loudspeaker position, the imaginary component is essentially zero. Below approximately 400 Hz, the imaginary component is below -40 dB for virtual sources in the front horizontal plane within the span of the loudspeakers. For the loudspeaker rear mirror position, the imaginary component becomes increasingly large as frequency increases, indicating some amount of frequency-dependent delay is eventually required in order to reproduce the rear source image, even if the magnitude of the ipsilateral loudspeaker still dominates in these situations.

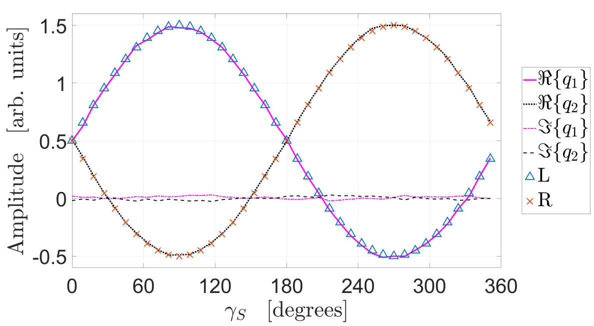

Fig. 1. Real and imaginary components of q_1, q_2 for $f \approx 52$ Hz versus sine law signals, L and R, as functions of virtual source angle γ_S.

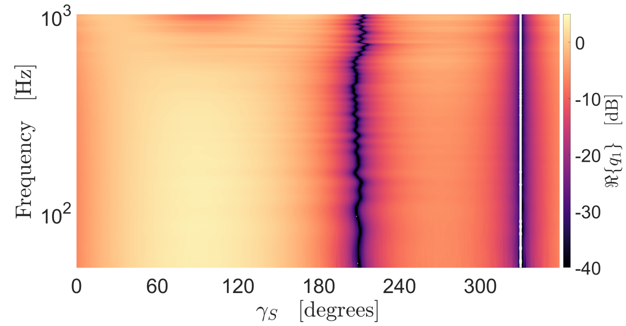

Fig. 2. Real component of q_1 as a function of virtual source angle γ_S and frequency up to 1000 Hz.

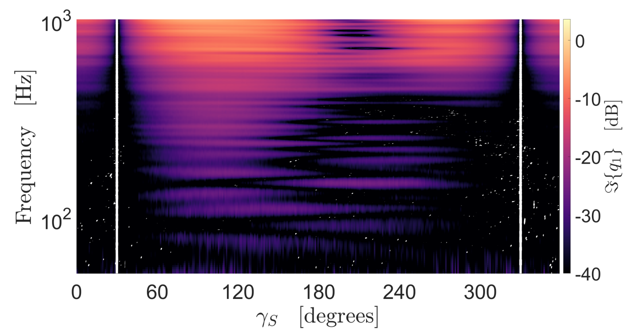

Fig. 3. Imaginary component of q_1 as a function of virtual source angle γ_S and frequency up to 1000 Hz.

8. CONCLUSION

In this paper, the analytical solution for the two loudspeaker CTC source strengths was presented assuming plane wave sources. It was shown for low frequency reproduction that this solution amounts to a scaled ratio of phase differences between the transfer functions used for the radiation matrix and the desired binaural signal.

It was shown using the rigid sphere head model that the low frequency approximation of the CTC source strengths considering an incident plane wave virtual source is equivalent to the stereo sine law when both the radiation matrix elements and desired binaural signal are commonly derived from the far field rigid sphere HRTF.

Due to the established accuracy of the rigid sphere head model compared to human HRTFs in the low frequencies, it was predicted that the low frequency source strengths for a more realistic far field HRTF, i.e., mannequin or human, would also converge to the stereo sine law in the low frequency limit. To test this hypothesis, publicly available KEMAR HRTFs supplied by TU-Berlin were used to create the source strengths. It was shown that for low frequencies, the real component of the solution aligns well with the stereo sine law while the imaginary component was comparatively small, confirming the proposed hypothesis.

These simulation results suggest CTC systems can be approximated at low frequencies to a desired accuracy with only amplitude panning, under the stated assumptions. The validity of this solution breaks down with increasing frequency as the imaginary component of the solution grows in magnitude, for close proximity sources and for mismatches in transfer functions.

9. REFERENCES

[1] B. B. Bauer, "Stereophonic earphones and binaural loudspeakers," *J. Audio Eng. Soc*, vol. 9, no. 2, pp. 148–151, 1961.

[2] M. F. Simón-Gálvez, T. Takeuchi, and F. M. Fazi, "Low-complexity, listener's position-adaptive binaural reproduction over a loudspeaker array," *Acta Acustica united with Acustica*, vol. 103, 2017.

[3] F. M. Fazi and E. C. Hamdan, "Stage compression in transaural audio," in *Audio Engineering Society Convention 144*, May 2018.

[4] P. A. Nelson, H. Hamada, and S. J. Elliott, "Adaptive inverse filters for stereophonic sound reproduction," *IEEE Transactions on Signal Processing*, vol. 40, no. 7, pp. 1621–1632, July 1992.

[5] J. Bauck and D. H. Cooper, "Generalized transaural stereo and applications," *J. Audio Eng. Soc*, vol. 44, no. 9, pp. 683–705, 1996.

[6] T. Takeuchi and P. A. Nelson, "Optimal source distribution for binaural synthesis over loudspeakers," *The Journal of the Acoustical Society of America*, vol. 112, no. 6, pp. 2786–2797, Dec. 2002.

[7] Y. Huang, J. Benesty, and J. Chen, "On crosstalk cancellation and equalization with multiple loudspeakers for 3-d sound reproduction," *IEEE Signal Processing Letters*, vol. 14, no. 10, pp. 649–652, Oct. 2007.

[8] S. Mori, M. Inoue, K. Matsui, and S. Adachi, "Stable inversion design for binaural reproduction over loudspeakers," in *2017 IEEE Conference on Control Technology and Applications (CCTA)*, Aug. 2017, pp. 1213–1217.

[9] P. A. Nelson and S. H. Yoon, "Estimation of acoustic source strength by inverse methods: Part i, conditioning of the inverse problem," *Journal of Sound and Vibration*, vol. 233, no. 4, pp. 639 – 664, 2000.

[10] F. M. Fazi and P. A. Nelson, "The ill-conditioning problem in sound field reconstruction," in *Audio Engineering Society Convention 123*, Oct. 2007.

[11] S. J. Elliott, J. Cheer, J. Choi, and Y. Kim, "Robustness and regularization of personal audio systems," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 7, pp. 2123–2133, Sept. 2012.

[12] T. Takeuchi and P. A. Nelson, "Extension of the optimal source distribution for binaural sound reproduction," *Acta Acustica united with Acustica*, vol. 94, no. 6, pp. 981–987, Nov. 2008.

[13] G. F. Kuhn, "Model for the interaural time differences in the azimuthal plane," *The Journal of the Acoustical Society of America*, vol. 62, no. 1, pp. 157–167, July 1977.

[14] R. O. Duda and W. L. Martens, "Range dependence of the response of a spherical head model," *The Journal of the Acoustical Society of America*, vol. 104, no. 5, pp. 3048–3058, 1998.

[15] M. F. Simón-Gálvez and F. M. Fazi, "Loudspeaker arrays for transaural reproduction," *The 22nd International Congress on Sound and Vibration*, 2015.

[16] ISO, "Quantities and units – part 2: Mathematical signs and symbols to be used in the natural sciences and technology," Standard, International Organization for Standardisation, Geneva, CH, Dec. 2009.

[17] B. Rafaely, *Fundamentals of Spherical Array Processing*, Springer-Verlag GmbH, 2015.

[18] B. Xie, *Head-Related Transfer Function and Virtual Auditory Display*, J Ross Publishing, 2013.

[19] D. Menzies, M. F. Simón-Gálvez, and F. M. Fazi, "A low-frequency panning method with compensation for head rotation," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 2, pp. 304–317, Feb. 2018.

[20] H. Wierstorf, M. Geier, and S. Spors, "A free database of head related impulse response measurements in the horizontal plane with multiple distances," in *Audio Engineering Society Convention 130*, May 2011.