Spectral manipulation improves elevation perception with non-individualized head-related transfer functions

Vani G. Rajendran
Department of Biomedical Sciences, City University of Hong Kong, 31 To Yuen Street, Kowloon Tong, Hong Kong
vrajendr@cityu.edu.hk

Hannes Gamper
Audio and Acoustics Research Group, Microsoft Research, 14865 Northeast 36th Street, Redmond, Washington 98052, USA
hannes.gamper@microsoft.com

Abstract: Spatially rendering sounds using head-related transfer functions (HRTFs) is an important part of creating immersive audio experiences for virtual reality applications. However, elevation perception remains challenging when generic, non-personalized HRTFs are used. This study investigated whether digital audio effects applied to a generic set of HRTFs could improve sound localization in the vertical plane. Several of the tested effects significantly improved elevation judgment, and trial-by-trial variability in spectral energy between 2 and 10 kHz correlated strongly with perceived elevation. Digital audio effects may therefore be a promising strategy to improve elevation perception where personalized HRTFs are not available.

[CCC]
Date Received: November 20, 2018 **Date Accepted:** February 21, 2019

1. Introduction

Head-related transfer functions (HRTFs) describe the spectral filtering that occurs due to the unique structure of a listener's head, torso, and pinnae, and can be measured at each of a listener's ears. Spatially rendering audio by filtering sound delivered to each ear through individual HRTFs confers the ability to localize sounds over headphones almost as accurately as in the free field (Wightman and Kistler, 1989a,b). However, individual HRTFs are impractical to acquire for large numbers of people, so a generic HRTF set (e.g., measured from a dummy head) is often used. While non-individual HRTFs can be effective in externalizing sounds (Seeber and Fastl, 2003), sound localization ability, particularly in elevation, is severely diminished (Begault *et al.*, 2001).

Unlike azimuth perception, where the auditory system has access to binaural cues such as interaural time and level differences, elevation perception relies primarily on monaural spectral cues that vary as a function of elevation in a highly individual manner (Blauert, 1997). It has been shown that a listener's anthropometric features can be related to elevation cues to personalize parametric HRTF models (Algazi *et al.*, 2002; Zotkin *et al.*, 2003; Satarzadeh *et al.*, 2007; Spagnol *et al.*, 2013), though in practice, a listener's anthropometry is typically unknown. The spectral peaks and notches that influence perceived elevation in free-field listening conditions (Shaw and Teranishi, 1968; Bloom, 1977; Roffler and Butler, 1968) have also been shown to affect elevation perception in spatial rendering systems including wave-field synthesis (Lopez *et al.*, 2010) and amplitude-panning (Pulkki, 2001). We hypothesized that adding physiologically inspired digital audio effects to generic HRTFs may improve elevation perception in spatially rendered audio delivered over headphones. Through a three-alternative forced choice task where subjects reported whether broadband noise bursts came from above, below, or level with the horizon, we identified four audio effects that significantly improved elevation perception. Additionally, spectral energy in the 2–10 kHz range for our sounds correlated strongly with perceived elevation. These findings suggest that spectral manipulation may provide a simple and practical way to improve elevation perception in audio that is spatially rendered using generic HRTFs.

2. Methods

Fifteen paid participants took part in the study (age 22–49, 3 females, 14 right-handed), including the two authors. Subjects were seated in a quiet room in front of a laptop and were instructed to keep their head level and facing forward at all times. They were told that sounds presented over the headphones would come from the right hand side, and that they were to report whether sounds originated from above the horizon (UP), below the horizon (DN), or level with the horizon (MD, defined as a plane through the ears) by pressing the up, middle, or down keys on the number pad of a keyboard. 24 trials of each of the 54 experimental conditions (2 sound types $\times$ 3 elevations $\times$ 9 effects) were randomly interleaved and presented over 8 blocks to listeners. Each block lasted ~3–4 min depending on the subject's pace, and subjects were permitted to take a short break between blocks. The experimenter remained in the room at all times.

Stimuli were 250 ms in duration and consisted of five identical broadband noise bursts 30 ms in duration (5 ms cosine onset and offset ramps) and separated by 20 ms silent gaps. Two sound types were tested: white noise and pink (1/f frequency spectrum) noise, and stimuli were generated afresh for all trials and all participants to avoid any possible learning effects (cf. Agus *et al.*, 2010). Sounds were rendered at $+45°$, $0°$, and $-45°$ polar angle and $45°$ lateral angle (right hand side of the participant) in interaural coordinates (see Macpherson and Middlebrooks, 2002), at a distance of 1 m. All sounds were rendered using a custom HRTF engine and a generic HRTF set that was measured on a dummy head and interpolated using the spherical harmonics transform (Ahrens *et al.*, 2012). For an improved sense of sound source externalization, early reflections and reverberation were added to this baseline HRTF set. This served as the control condition, and all digital audio effects were applied on top of the control condition. Sounds were presented over AKG K702 headphones through a Fireface UCX soundcard without head tracking.

The eight effects tested are listed below with the relevant parameter values used. All effects were applied by altering the baseline HRTF set described above.

8 kHz peak filter: Numerous studies report that narrowband sounds centered around 8 kHz tend to be perceived as coming from overhead (Hebrank and Wright, 1974; Blauert, 1997). We hypothesized that applying an 8 kHz peak filter to sounds might therefore bias perception upwards. The 8 kHz peak filter had a center frequency (CF) of 8 kHz, a Q of 6, and a 10 dB gain.

8 kHz notch filter: Similarly, we hypothesized that selectively attenuating frequencies around 8 kHz might bias perception downwards (CF = 8 kHz, Q = 6, and gain = -10 dB).

7 kHz notch filter: There are data suggesting that the CF of the spectral notch that results from filtering by the pinnae tightly matches the perceived elevation of a sound source (Bloom, 1977). Based on these data, we expected a 7 kHz notch to bias perception downwards (CF = 7 kHz, Q = 6, gain = -10 dB).

13 kHz notch filter: According to the same study (Bloom, 1977), we expected that a 13 kHz spectral notch might bias perception upwards (CF = 13 kHz, Q = 6, gain = -10 dB).

Horizontal blur: By combining or blurring the localization features of nearby locations in an HRTF, the resulting average HRTF may become less specific to the measured dummy head and thereby generalize better across listeners. To create a horizontal blur, impulse responses from two points in the HRTF at $+/-$ $8°$ lateral angle to the target location were averaged in the time domain.

Vertical blur: Impulse responses from two points in the HRTF at $+/-$ $8°$ polar angle relative to the target location were averaged in the time domain.

Floor reflections: Though all conditions included early reflections and reverberation, a strong floor reflection (reflection coefficient = 1) was added to each virtual source.

Offset ears: The barn owl is extremely precise in localizing sounds in the median plane and does so by using binaural cues that result from a vertical offset between the two ears (Knudsen and Konishi, 1979). We simulated the effect of vertically offset ears by simulating a displacement of the right and left receiver positions by a polar angle of $8°$ above and below the horizontal axis, respectively (Gamper *et al.*, 2015).

3. Results

3.1 8 kHz peak filter, 7 kHz notch filter, and horizontal blur directionally bias perception

The audio effects tested could either bias perception upwards or downwards, or change the overall accuracy at all elevations tested. The mean directional bias was quantified by assigning UP responses a value of $+1$, MD responses a value of 0, and DN responses a value of -1, and then calculating the average across all responses within a condition. A value close to 1 would therefore correspond to a sound being perceived consistently above the horizon. Figure 1(A) shows the directional bias produced by each audio effect, pooled across sound type (white or pink noise) and elevation (−45°, 0°, +45°). If elevation perception was accurate, pooling across elevations should result in a directional bias of 0. Each boxplot therefore represents directional bias across the 15 subjects, and the large spread in the data reflects the inter-subject variability in elevation perception. Three effects show a clear directional bias relative to control, which itself shows a slight positive directional bias. The 8 kHz peak filter created a significant upward bias in elevation perception, while the 7 kHz notch filter and horizontal blur effects significantly biased perception downwards relative to control ($p < 0.001$, $N = 15$ subjects, Wilcoxon paired signed-rank test, Bonferroni corrected). The directional bias for these three effects was unanimous across subjects.

3.2 Vertical blur effect improves accuracy at all elevations tested

Accuracy was calculated as the total percent correct based on the target sound source elevation determined by the HRTF. Figure 1(B) shows accuracy for all subjects and all conditions, again pooled across sound types and elevations. Since this was a three-alternative forced choice task (UP, MD, DN), chance performance would be around 33%, which is roughly the median accuracy in the control condition (34%). Note that a directional bias would not necessarily result in a systematic change in accuracy; for example, an effect that biases perception at all elevations upward would improve accuracy for the +45° elevation but detract from it at the other two elevations. The only effect that produced a significant improvement in accuracy relative to control was the vertical blur effect ($p < 0.005$, $N = 15$ subjects, Wilcoxon paired signed-rank test, Bonferroni corrected).

3.3 Combining the best effects at each elevation substantially improves localization accuracy

Table 1 breaks down accuracy by elevation, allowing a comparison of mean accuracy between all tested effects, the control, the best audio effect for that elevation, and the improvement in percentage points of the best audio effect relative to control. The same four effects identified earlier appear again: the horizontal blur, 7 kHz notch filter, 8 kHz peak filter, and vertical blur effects result in the highest accuracy for −45°, 0°, +45°, and all elevations (ALL), respectively. These observations hold even when data are split by sound type (white or pink noise). However, some overall differences in

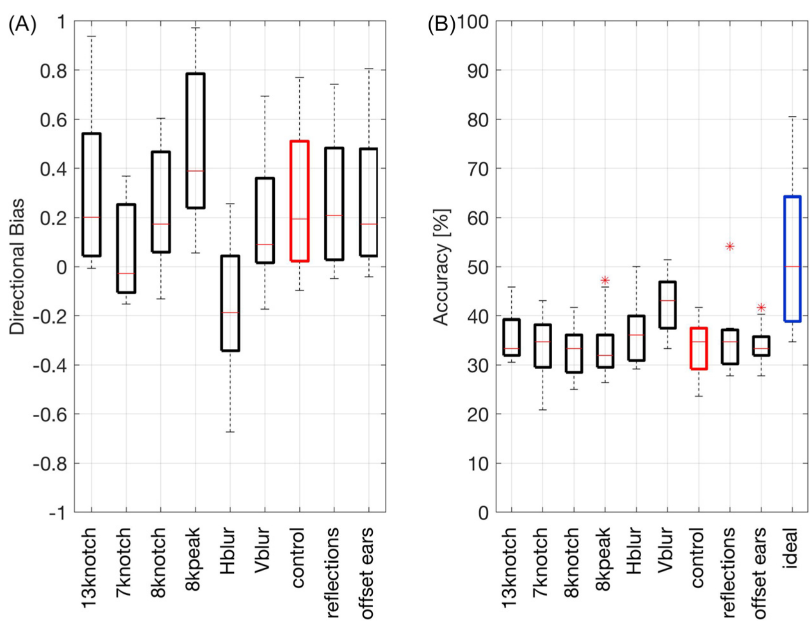

Fig. 1. (Color online) (A) Directional bias across subjects observed for each audio effect, pooled across sound type and elevation. UP, MD, and DN responses were assigned a value of +1, 0, and −1, respectively, and directional bias was calculated as the average across all responses in a condition. A positive directional bias would indicate that most stimuli were perceived from above the horizon. The 8 kHz peak, 7 kHz notch, and horizontal blur effects show a significant directional bias relative to the control condition ($p < 0.001$, $N = 15$ subjects, Wilcoxon paired signed-rank test, Bonferroni corrected). (B) Accuracy [% correct] of responses across subjects for each audio effect, pooled across sound type and elevation. The rightmost boxplot shows that an "ideal" condition that combined the most accurate effects at each elevation would have achieved a median accuracy of 51%.

Table 1. Task accuracy broken down by elevation and pooled across sound type (white or pink noise). Columns show the accuracy as a percentage for the average across all conditions tested, the control condition, the best (highest accuracy) condition at each elevation, and the improvement in percentage points of the best condition over the control. Combining the best effects at each elevation would result in a median accuracy of 51% [see "ideal" in Fig. 1(B)].

Elevation	Avg. [%]	Control [%]	Best [%]	Improv. [%]	Best condition
DN	20.8	16.7	46.4	29.7	Hblur
MD	45.7	43.8	54.3	10.6	7knotch
UP	38	40.3	57.9	17.6	8kpeak
ALL	34.8	33.6	40.9	7.4	Vblur

accuracy between white and pink noise conditions can be seen in Tables 2 and 3. Combining the best effect at each elevation substantially improves median accuracy from 34% in the control condition to 51% [rightmost boxplot in Fig. 1(B)].

3.4 Spectral energy between 2 and 10 kHz correlates with perceived elevation on a trial-by-trial basis

Since stimuli were generated afresh for each trial and for each participant, the 19 440 unique trials collected (2 sound types $\times$ 3 elevations $\times$ 9 effects $\times$ 24 repeats $\times$ 15 subjects) presented an opportunity to explore the effect of spectral content on a trial-by-trial basis. For each trial, the spectral energy was computed in 1/3-octave frequency bands from 16 Hz to 24 kHz. Figure 2(A) shows the average spectrum at the right ear for correct and incorrect trials by elevation and sound type. The average spectra for correct and incorrect signals consistently diverge between 2 and 10 kHz; higher energy in this range is observed for correct trials at UP elevations and incorrect trials at DN elevations. Figure 2(B) reveals a strong correlation between summed spectral energy between 2 and 10 kHz and perceived elevation (white noise: $R = 0.84$, $p < 0.01$; pink noise: $R = 0.91$, $p < 0.01$, Pearson correlation).

4. Discussion

In this study, we found that adding simple digital audio effects to broadband noise bursts before filtering them through a pair of generic HRTFs can significantly improve elevation perception. The effects identified include an 8 kHz peak filter (biased perception upwards), 7 kHz notch filter and horizontal blur effects (biased perception downwards), and vertical blur (improved accuracy at all elevations tested). Furthermore, trial-by-trial variability in spectral energy between 2 and 10 kHz correlated strongly with perceived elevation.

While we have demonstrated that digital audio effects can improve elevation perception for broadband sounds, whether they are as effective for natural sounds such as speech, and at low enough gains to avoid perceptible distortions, remain open questions. It is worth noting, however, that spectral peaks and notches are features that are naturally introduced by the pinnae under free-field listening conditions and are readily interpreted by the brain as an elevation cue (Bloom, 1977; Blauert, 1997; Rice *et al.*, 1992; Middlebrooks, 1999; Langendijk and Bronkhorst, 2002; Macpherson and Sabin, 2013). For example, a peak around 8 kHz has been reported repeatedly as an "above" cue in free-field listening conditions (Hebrank and Wright, 1974; Blauert, 1997), and our findings suggest that the same applies in virtual acoustic space. Similarly, a spectral notch around 7 kHz is also associated with elevations below the horizon (Bloom, 1977; Hofman and Opstal, 2002) and biased perception downward in our study. Future work is required to determine whether these consistencies are an indication of certain critical spectral features that may alone be sufficient to bias perceived elevation. Additional considerations include whether these effects can be combined, whether they

Table 2. Task accuracy broken down by elevation, but for white noise only. See Table 1.

Elevation	Avg [%]	Control [%]	Best [%]	Improv [%]	Best condition
DN	12.7	9.2	38.1	28.9	Hblur
MD	40.8	39.4	52.5	13.1	7knotch
UP	49.7	53.3	67.2	13.9	8kpeak
ALL	34.4	34.0	39.4	5.5	Vblur

Table 3. Task accuracy broken down by elevation, but for pink noise only. See Table 1.

Elevation	Avg [%]	Control [%]	Best [%]	Improv [%]	Best condition
DN	29.0	24.2	54.7	30.6	Hblur
MD	50.6	48.1	56.1	8.1	7knotch
UP	26.2	27.2	48.6	21.4	8kpeak
ALL	35.2	33.1	42.4	9.3	Vblur

are as effective at different azimuths, and how they might be parametrized (Iida *et al.*, 2007; Yao *et al.*, 2018).

The strong downward bias produced by the horizontal blur condition may be due to a comb filter-like effect introduced into the frequency spectrum by the time-domain averaging of impulse responses with slight timing differences between the two lateral HRTF locations. Even under normal listening conditions, reflections from the torso can act as a comb filter whose pattern of notches varies with elevation (Algazi *et al.*, 2001). The particular comb filter produced by our stimulus may have mapped to an elevation below the horizon. In contrast, vertical blur had the intended effect of improving accuracy at all elevations. The offset ears condition, though it introduced a binaural cue for the judgment of elevation, did not improve elevation perception. Since this binaural elevation cue is not present under normal listening conditions, it is possible that listeners were unable to interpret this additional information. Given that the auditory system can learn and maintain multiple spatial auditory maps (Kumpik *et al.*, 2010), this effect may be worth testing in conjunction with a training phase.

While combining the top performing effects at each elevation improved median accuracy from 34% to 51%, there is a large degree of inter-subject variability in the efficacy of the so-called best effects, with one subject still performing around the 33% chance level and another at 78% accuracy. In practice, it may be possible to attain further improvements in accuracy by personalizing the audio effects applied at each elevation through a short calibration phase.

The identification of digital audio effects that generalize across listeners represents an alternative but complementary approach to the main thrust of work in the field, which is on HRTF personalization. HRTF personalization methods are often based on anthropometric measurements (Zotkin *et al.*, 2003; Satarzadeh *et al.*, 2007; Bilinski *et al.*, 2014; Tashev, 2014), which can be used to select a closely matching

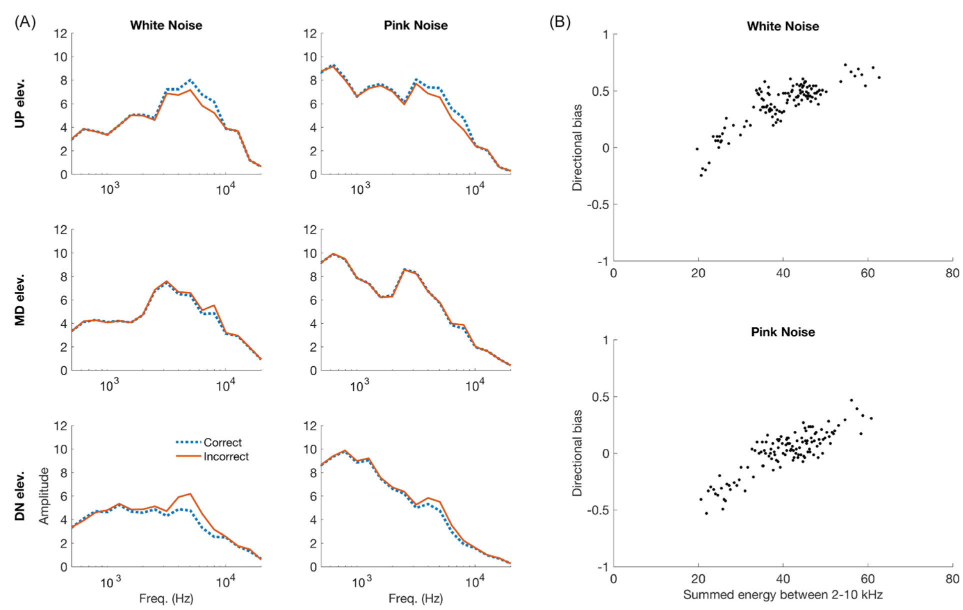

Fig. 2. (Color online) (A) Each panel shows the average frequency spectrum of individual trials for each sound type (columns) and each elevation (rows). The data are split by correct (dotted lines) and incorrect (solid lines) responses. Note the divergence in the average frequency spectra between correct and incorrect responses in the 2–10 kHz range. (B) The summed energy between 2 and 10 kHz plotted against mean directional bias. There is a strong positive correlation between spectral energy in this range and perceived elevation above the horizon for both sound types tested (white noise: $R = 0.84$, $p < 0.01$; pink noise: $R = 0.91$, $p < 0.01$, Pearson correlation).

HRTF from a database (Seeber and Fastl, 2003). Computational approaches to HRTF personalization include using 3D spherical transforms (Politis, *et al.*, 2016) and artificial neural networks (Hu *et al.*, 2008). HRTF synthesis is also an approach commonly employed (Rund and Saturka, 2012; Brown and Duda, 1998; Algazi *et al.*, 2002). Our approach was to identify general features that influence elevation perception. Other studies that take this approach include using machine learning to discover elevation cues in individualized HRTFs (Thuillier *et al.*, 2018), and parametrizing spectral peaks and notches found in individualized HRTFs to create the illusion of elevation in binaural room impulse responses (Yao *et al.*, 2018).

5. Conclusions

Digital audio effects may be an effective and practical way to improve elevation perception in spatially rendered audio using generic HRTFs. The effects identified here (8 kHz peak and 7 kHz notch filters, horizontal and vertical blur, and total spectral energy between 2 and 10 kHz) robustly affected elevation judgement and, in their ideal combination, improved accuracy by 17 percentage points in a 3-alternative forced choice sound localization task. These findings may be applied to improving the audio experience in immersive mixed reality applications and may inform future investigations in to the neural bases of perceived elevation (Ahveninen *et al.*, 2014). Investigating how these effects might be incorporated into a parametric HRTF model that allows personalization given a listener's anthropometric features may be a promising avenue for future work.

Acknowledgments

This work was carried out while V.G.R. was an intern at Microsoft Research, Cambridge, UK.

References and links

Agus, T. R., Thorpe, S. J., and Pressnitzer, D. (**2010**). "Rapid formation of robust auditory memories: Insights from noise," Neuron **66**(4), 610–618.

Ahrens, J., Thomas, M. R., and Tashev, I. (**2012**). "HRTF magnitude modeling using a non-regularized least-squares fit of spherical harmonics coefficients on incomplete data," in *Proceedings of The 2012 Asia Pacific Signal and Information Processing Association Annual Summit and Conference*, IEEE, pp. 1–5.

Ahveninen, J., Kopčo, N., and Jääskeläinen, I. P. (**2014**). "Psychophysics and neuronal bases of sound localization in humans," Hear. Res. **307**, 86–97.

Algazi, V. R., Avendano, C., and Duda, R. O. (**2001**). "Elevation localization and head-related transfer function analysis at low frequencies," J. Acoust. Soc. Am. **109**(3), 1110–1122.

Algazi, V. R., Duda, R. O., and Thompson, D. M. (**2002**). "The use of head-and-torso models for improved spatial sound synthesis," in *Audio Engineering Society Convention 113*, Audio Engineering Society.

Begault, D. R., Wenzel, E. M., and Anderson, M. R. (**2001**). "Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source," J. Audio Eng. Soc. **49**(10), 904–916.

Bilinski, P., Ahrens, J., Thomas, M. R., Tashev, I. J., and Platt, J. C. (**2014**). "HRTF magnitude synthesis via sparse representation of anthropometric features," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, pp. 4468–4472.

Blauert, J. (**1997**). *Spatial Hearing: The Psychophysics of Human Sound Localization* (MIT press, Cambridge, MA).

Bloom, P. J. (**1977**). "Creating source elevation illusions by spectral manipulation," J. Audio Eng. Soc. **25**(9), 560–565.

Brown, C. P., and Duda, R. O. (**1998**). "A structural model for binaural sound synthesis," IEEE Trans. Speech Audio Process. **6**(5), 476–488.

Gamper, H., Thomas, M. R. P., and Tashev, I. J. (**2015**). "Anthropometric parameterisation of a spherical scatterer ITD model with arbitrary ear angles," in *2015 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, IEEE, pp. 1–5.

Hebrank, J., and Wright, D. (**1974**). "Spectral cues used in the localization of sound sources on the median plane," J. Acoust. Soc. Am. **56**(6), 1829–1834.

Hofman, P. M., and Van Opstal, A. J. (**2002**). "Bayesian reconstruction of sound localization cues from responses to random spectra," Biol. Cybern. **86**(4), 305–316.

Hu, H., Zhou, L., Ma, H., and Wu, Z. (**2008**). "HRTF personalization based on artificial neural network in individual virtual auditory space," Appl. Acoust. **69**(2), 163–172.

Iida, K., Itoh, M., Itagaki, A., and Morimoto, M. (**2007**). "Median plane localization using a parametric model of the head-related transfer function based on spectral cues," Appl. Acoust. **68**(8), 835–850.

Knudsen, E. I., and Konishi, M. (**1979**). "Mechanisms of sound localization in the barn owl (*Tyto alba*)," J. Comp. Physiol. **133**(1), 13–21.

Kumpik, D. P., Kacelnik, O., and King, A. J. (**2010**). "Adaptive reweighting of auditory localization cues in response to chronic unilateral earplugging in humans," J. Neurosci. **30**(14), 4883–4894.

Langendijk, E. H., and Bronkhorst, A. W. (**2002**). "Contribution of spectral cues to human sound localization," J. Acoust. Soc. Am. **112**(4), 1583–1596.

Lopez, J. J., Cobos, M., and Pueo, B. (**2010**). "Elevation in wave-field synthesis using HRTF cues," Acta Acust. Acust. **96**(2), 340–350.

Macpherson, E. A., and Middlebrooks, J. C. (**2002**). "Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited," J. Acoust. Soc. Am. **111**(5), 2219–2236.

Macpherson, E. A., and Sabin, A. T. (**2013**). "Vertical-plane sound localization with distorted spectral cues," Hear. Res. **306**, 76–92.

Middlebrooks, J. C. (**1999**). "Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency," J. Acoust. Soc. Am. **106**(3), 1493–1510.

Politis, A., Thomas, M. R., Gamper, H., and Tashev, I. J. (**2016**). "Applications of 3D spherical transforms to personalization of head-related transfer functions," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, pp. 306–310.

Pulkki, V. (**2001**). "Localization of amplitude-panned virtual sources II: Two-and three-dimensional panning," J. Audio Eng. Soc. **49**(9), 753–767.

Rice, J. J., May, B. J., Spirou, G. A., and Young, E. D. (**1992**). "Pinna-based spectral cues for sound localization in cat," Hear. Res. **58**(2), 132–152.

Roffler, S. K., and Butler, R. A. (**1968**). "Factors that influence the localization of sound in the vertical plane," J. Acoust. Soc. Am. **43**(6), 1255–1259.

Rund, F., and Saturka, F. (**2012**). "Alternatives to HRTF measurement," in *2012 35th International Conference on Telecommunications and Signal Processing (TSP)*, IEEE, pp. 648–652.

Satarzadeh, P., Algazi, V. R., and Duda, R. O. (**2007**). "Physical and filter pinna models based on anthropometry," in *Audio Engineering Society Convention 122*, Audio Engineering Society.

Seeber, B. U., and Fastl, H. (**2003**). "Subjective selection of non-individual head-related transfer functions," Georgia Institute of Technology.

Shaw, E. A. G., and Teranishi, R. (**1968**). "Sound pressure generated in an external-ear replica and real human ears by a nearby point source," J. Acoust. Soc. Am. **44**(1), 240–249.

Spagnol, S., Geronazzo, M., and Avanzini, F. (**2013**). "On the relation between pinna reflection patterns and head-related transfer function features," IEEE Trans. Audio Speech Lang. Process. **21**(3), 508–519.

Tashev, I. (**2014**). "HRTF phase synthesis via sparse representation of anthropometric features," in *2014 Information Theory and Applications Workshop (ITA)*, IEEE, pp. 1–5.

Thuillier, E., Gamper, H., and Tashev, I. J. (**2018**). "Spatial audio feature discovery with convolutional neural networks," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, pp. 6797–6801.

Wightman, F. L., and Kistler, D. J. (**1989a**). "Headphone simulation of free-field listening. I: Stimulus synthesis," J. Acoust. Soc. Am. **85**(2), 858–867.

Wightman, F. L., and Kistler, D. J. (**1989b**). "Headphone simulation of free-field listening. II: Psychophysical validation," J. Acoust. Soc. Am. **85**(2), 868–878.

Yao, D., Li, J., Xia, R., and Yang, F. (**2018**). "A parametric method for elevation control," in *2018 16th International Workshop on Acoustic Signal Enhancement (IWAENC)*, IEEE, pp. 181–185.

Zotkin, D. Y. N., Hwang, J., Duraiswaini, R., and Davis, L. S. (**2003**). "HRTF personalization using anthropometric measurements," in *2003 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, IEEE Cat. No. 03TH8684, IEEE, pp. 157–160.