OPEN ACCESS Freely available online



Audio Engineering Society

Convention Paper 9870

Presented at the 143rd Convention
2017 October 18–21, New York, NY, USA

This convention paper was selected based on a submitted abstract and 750-word precis that have been peer reviewed by at least two qualified anonymous reviewers. The complete manuscript was not peer reviewed. This convention paper has been reproduced from the author's advance manuscript without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. This paper is available in the AES E-Library (http://www.aes.org/e-lib), all rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

Apparent Sound Source De-elevation Using Digital Filters Based On Human Sound Localization

Adrian Celestinos[1], Elisabeth McMullin[1], Ritesh Banka[1], William DeCanio[1], and Allan Devantier[1]

[1] *Samsung Research America*

Correspondence should be addressed to Adrian Celestinos (a.celestinos@samsung.com)

ABSTRACT

This study presents the possibility of creating an apparent sound source elevated or de-elevated from its current physical location. When loudspeakers need to be placed in different locations other than the ideal placement, digital filters are created and connected in the audio chain to either elevate or de-elevate the perceived sound from its physical location. The filters are based on Head-Related Transfer Functions (HRTF) measured in human subjects. The filters relate to an average from individual de-elevation filters, representing generalized transfer functions of humans for sources in the frontal median plane. Results showed that a universal de-elevation filter for an actual source at 25°can create an apparent source de-elevated by about 13° for male speech and by 10° for female speech.

1 Introduction

In sound reproduction there are cases where the loudspeakers need to be displaced from the ideal location (e.g. on a TV screen the speakers may need to be placed either on top of the screen or below it). This setup would separate the sound source from the picture creating an undesirable effect. In order to deal with the idea of creating an apparent sound source elevated or de-elevated from its physical location, understanding the mechanism of human sound localization in the frontal median plane is of high importance. A listener is able to perceive the direction of a sound source because the sound, on its way to the ear drum is modified or filtered by diffractions and reflections from the human head, pinna and torso. Our hearing is able to "recognize" the filtering and thus determine direction to the source [1]. A Head-Related Transfer Function (HRTF) contains the directional information embedded in the transmission path from a sound source to the human ear. If we restrict this study to the case of only one sound source, then a starting point would be to look at the sound path from the source in free field to the ears of a human subject.

The cues contained in the HRTF (i.e. Interaural Time Difference (ITD), Interaural Level Difference (ILD) and spectral changes), help the listener to localize a sound event. Localization in the horizontal plane is based on the ITDs (time arrival to ears), ILDs (produced by head shadowing), and spectral changes caused by reflections and diffraction of the head, torso, and pinnae. Localization in the median plane is different than in the horizontal plane since the signals available at both ears are almost identical. Therefore, in the

median plane the cues for localization can be reduced to monaural spectral stimuli. The localization blur for changes in elevation of the sound source in the forward direction is approximately 17° (continuous speech by unfamiliar person) [2]. In the literature, Lopez et al. [3] have proposed a hybrid method to elevate dynamic sound sources in a wave field synthesis installation using a "universal" HRTF average from two different databases.

The goal in this paper is to analyze the characteristics of HRTFs corresponding to the median plane and produce a digital filter to convolve with the input signal to the displaced sound source (loudspeaker); thus the listener will perceive sound coming from the de-elevated direction. First, the methodology to create the de-elevation filters is outlined in Section 2; afterwards, the evaluation of de-elevation filters is delineated in Section 2.7. Results of the evaluation are presented in Section 3 and finally discussion and conclusions are given in Sections 4 and 5 respectively.

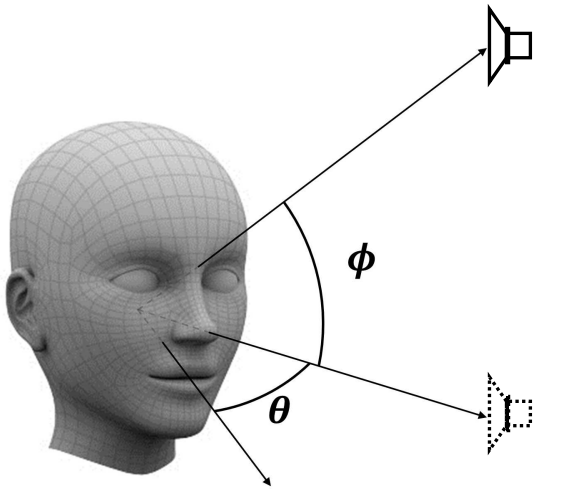


Fig. 1: Polar coordinate to describe HRTF, ϕ is the elevation angle and θ is the azimuth angle, adapted from [5].

2 Methods

In this section a detailed description of the methods utilized for analysis and construction of the de-elevation filters is presented.

2.1 Head-Related Transfer Functions

A Head-Related transfer function is a transfer function that, for certain angle of incidence, describes the sound transmission from a free field to a point in the ear canal of a human subject [1] (Fig.1). HRTFs are computed by Eq.(1) and Eq.(2), where P_1 corresponds to the sound pressure at the center position of the head (head absent), and P_2 refers to the sound pressure at the entrances of left and right blocked ear canals respectively.

$$HRTF_{Left\ ear}(\phi,\theta) = \frac{P_{2_{Left\ ear}}}{P_1}(\phi,\theta) \quad (1)$$

$$HRTF_{Right\ ear}(\phi,\theta) = \frac{P_{2_{Right\ ear}}}{P_1}(\phi,\theta) \quad (2)$$

IRCAM Database

There are a number of available HRTF databases. For this study it was important to analyze data measured from a large number of human subjects. The "Listen" database that was kindly made available by the IRCAM Institute for the scientific community was used for the purpose of analysis [4]. This collection of data contains Head-Related Impulse Responses (HRIR) in the range from -45° to 90°with an elevation angle resolution of 15°. The impulse responses are measured in the blocked entrance of the ear canal of each subject. Since the focus of this paper is the directional hearing in the median plane with the sound source in the forward direction we have extracted HRIRs from -45° to 75° elevation angle in the median plane only.

Audio Lab Database

In addition to the IRCAM database, HRTF measurements were performed on 14 human subjects and one dummy head in the anechoic chamber of Samsung Audio Lab in Valencia, CA. HRIRs were acquired in the median plane with the sound source in the forward direction having a resolution of 5°from $\phi = 0°$ to $\phi = 60°$ elevation. The impulse responses were recorded with miniature microphones inserted at the blocked ear canal entrance, as shown in Fig.11, following the methodology delineated by Møller et al. [1]. The impulse responses were computed utilizing the logarithmic sweep method detailed by Farina [6]. The measurement setup included a 2.5" full-range driver mounted in a sealed spherical enclosure. The sound source was clamped to an automated arc which was connected to a turntable



Fig. 2: HRTF Measurement setup at Samsung Audio Lab anechoic chamber.

controlled by a PC, as shown in Fig.2. Custom software (SAMSLab) controls the turntable which moves the sound source up and down and it acquires the impulse response with dedicated DSP audio hardware.

2.2 Pre-processing of HRIR

The raw HRIR data is truncated by multiplying it with an asymmetric window formed by two half-sided Blackman-Harris windows (Fig. 3). The final length of the HRIRs was 256 samples. Both impulse responses P_2 (left and right ears) and P_1, which includes the electro-acoustic chain, were transformed to the frequency domain using the discrete Fourier transform (DFT); then a complex division was performed in the

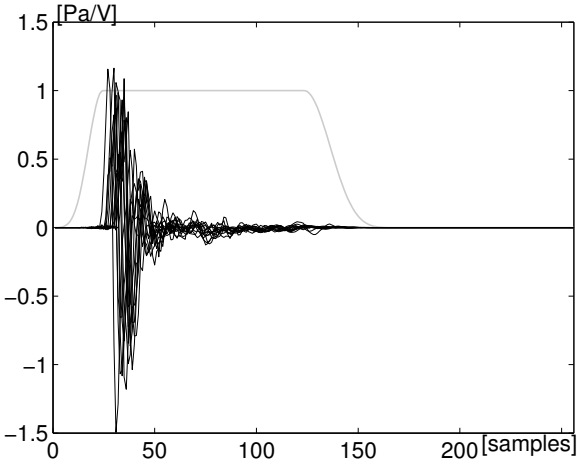


Fig. 3: Black lines, measured impulse responses for one subject. Gray line truncation window.

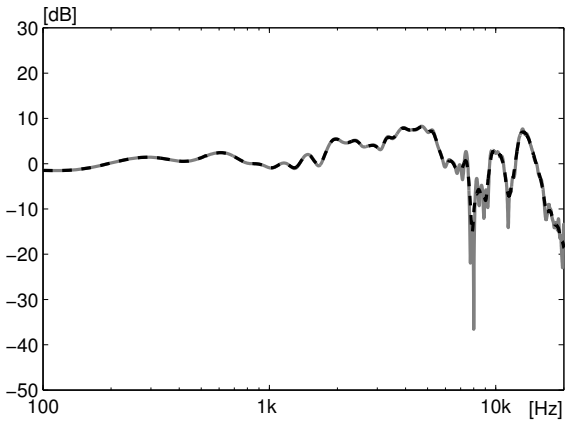


Fig. 4: Gray line, original HRTF, left ear subject 1016 at $\phi = 10^\circ$ elevation from IRCAM database. Black dashed line, smoothed version.

frequency domain to eliminate the effect of the electro-acoustic chain. By using the inverse Fourier transform, the transfer functions were returned to the time domain and low-pass filtered at 20 kHz. Finally, the DC component was removed from the impulse responses. Data from the IRCAM database was upsampled to fs = 48 kHz in order to match Audio Lab measurement data. Pre-processing and analysis was performed in MATLAB®.

2.3 Interpolation

For the IRCAM database, interpolation in the time domain has been performed to obtain one-degree resolution from -45°to 75°elevation in the median plane using only the HRIR corresponding to a zero azimuth angle (Fig. 5). The process was performed by using, shape-preserving piecewise cubic interpolation. The corresponding HRIRs of the 46 subjects' left ears were added to the right ears; thus a total of 92 impulse responses were computed. Once the required directional resolution was obtained, the HRIRs of both databases were transformed again to the frequency domain using an N=4800 point DFT. Since the final length of the HRIRs was 256 samples, the arrays were padded with trailing zeros to equal length N.

2.4 Smoothing

The HRTFs were smoothed using complex fractional octave smoothing. The amplitude and phase were

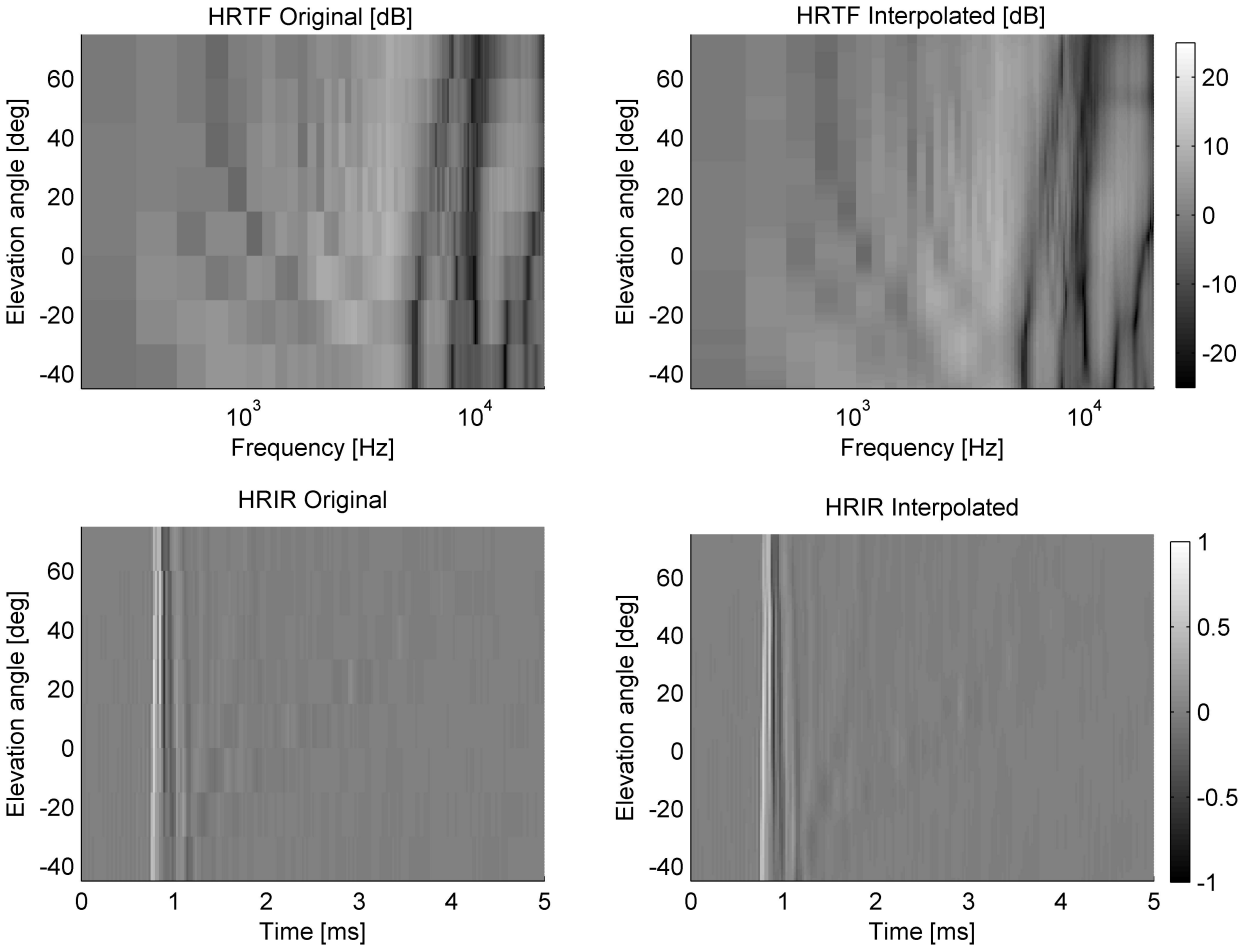


Fig. 5: Example of one subject's directional data, IRCAM database. Left column, original. Right column, interpolated. Upper row, HRTF. Lower row, HRIR.

smoothed separately with a $\frac{1}{12}$ octave bandwidth filter and a rectangular window (Fig. 4). By doing this procedure mainly the high Q notches were smoothed out from the HRTF, helping to produce a realizable filter.

2.5 HRTF Analysis

In Fig. 6, HRTF data from three subjects of each database is shown. The transfer functions are normalized to $\phi = 0^\circ$ elevation. As one can observe, there is a prominent peak around 1.25 kHz as elevation increases in both databases. Another obvious peak is around 6.5 kHz in all subjects, this observation was reported by E.A.G. Shaw in previous investigations [7]. It seems that there is another peak around 2.8-3.2 kHz in all six subjects, but it is not very clear. After observing the normalized HRTFs at zero degrees of elevation, one can infer that in order to de-elevate a sound source (e.g. at ϕ =25°), first, the effect of having the source at that angle has to be canceled, then the spectral cue corresponding to the apparent location has to be imposed. In the following section a method of creating the de-elevation filters is detailed.

2.6 De-elevation Filters

The de-elevation filter is defined by the complex division in the frequency domain as shown in Eq. (3). The numerator is the HRTF corresponding to the apparent

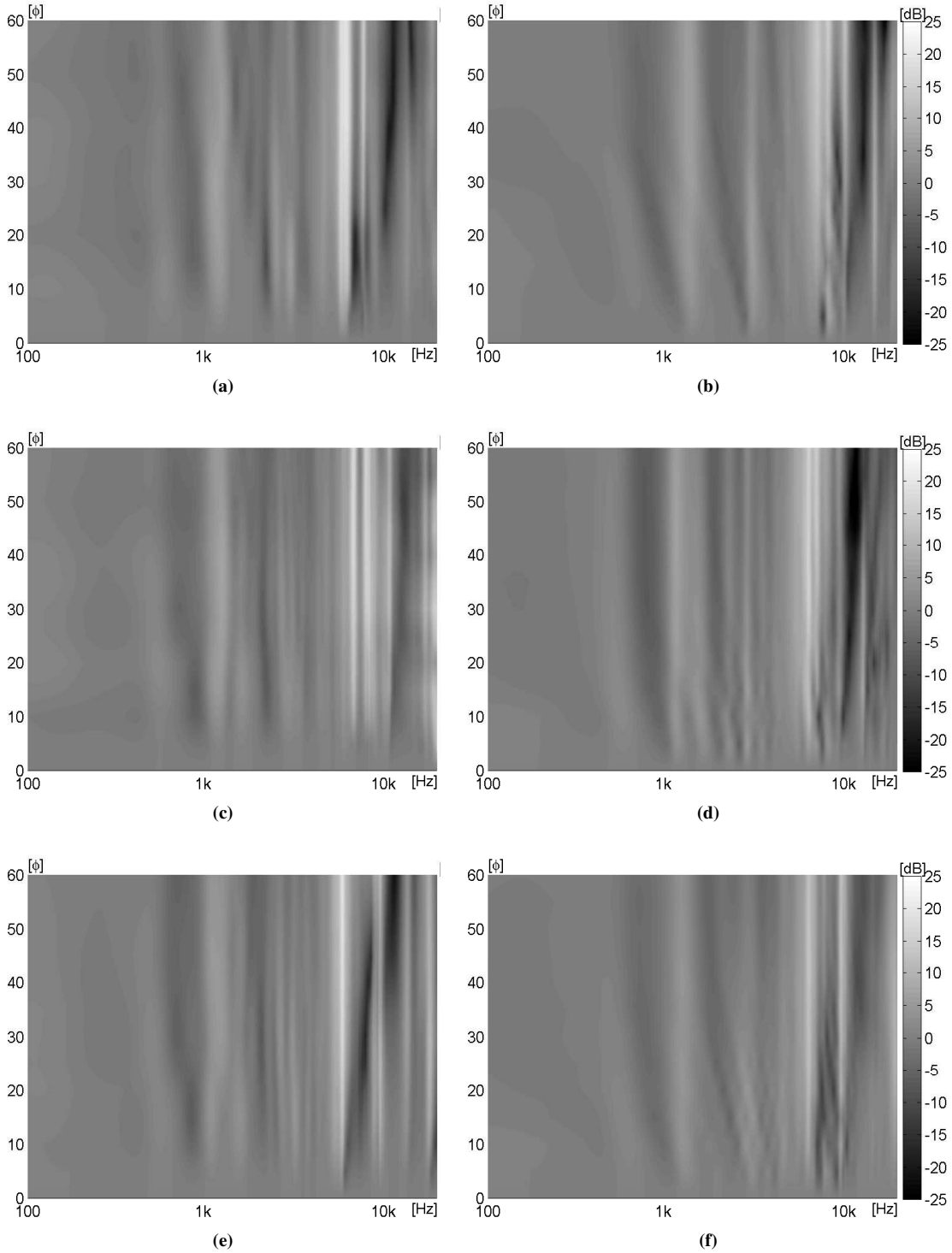


Fig. 6: HRTFs normalized to $\phi = 0^\circ$ elevation. Left column IRCAM database, plots (a), (c), and (e) are data from subjects 1018, 1020, and 1041. Right column Samsung Audio Lab database, plots (b), (d), and (f) data from subjects 3, 6, and 9.

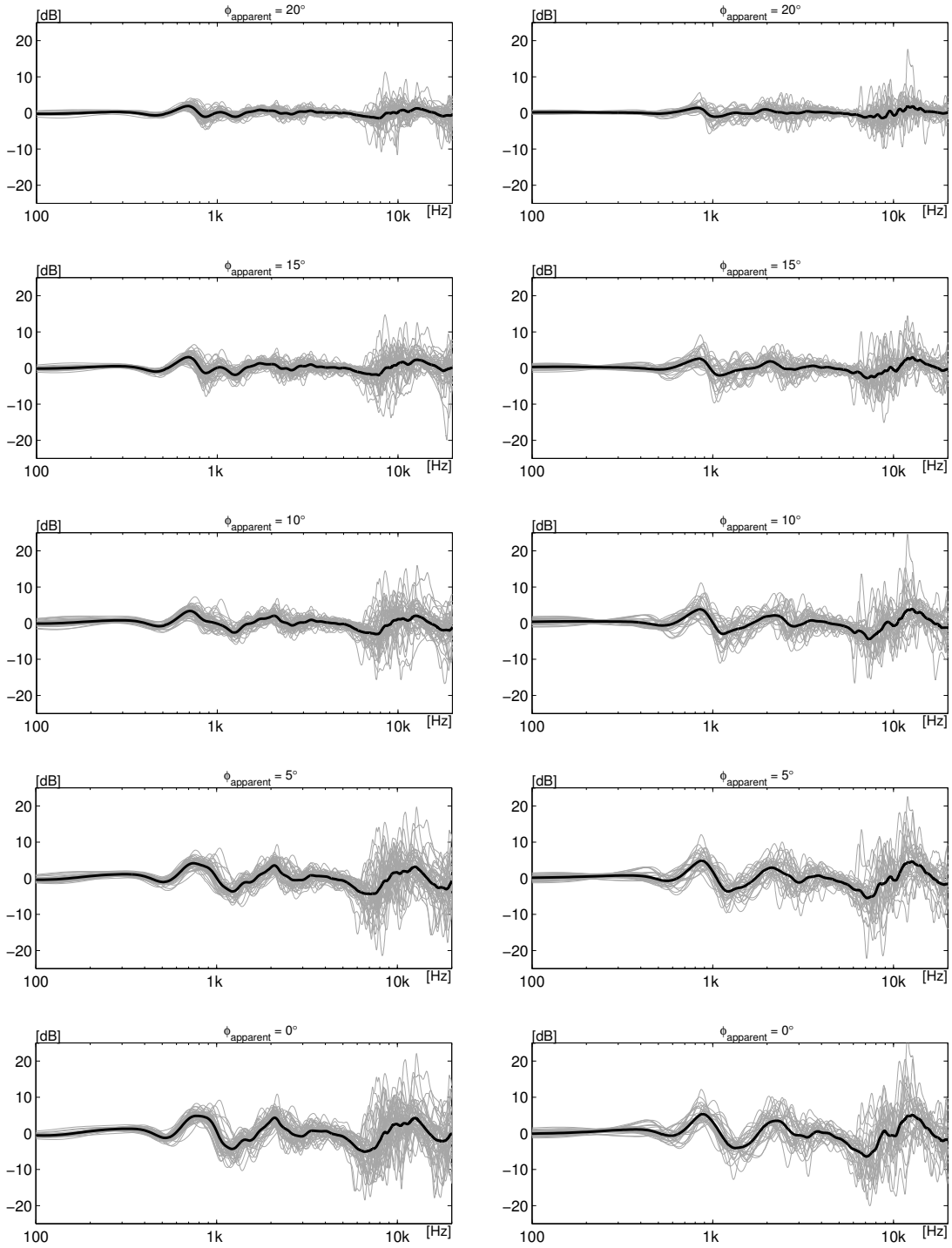


Fig. 7: Left column IRCAM average of de-elevation filters. Right column Audio Lab average of de-elevation filters. Light gray curves, individual de-elevation filters, black curves, averaged filter.

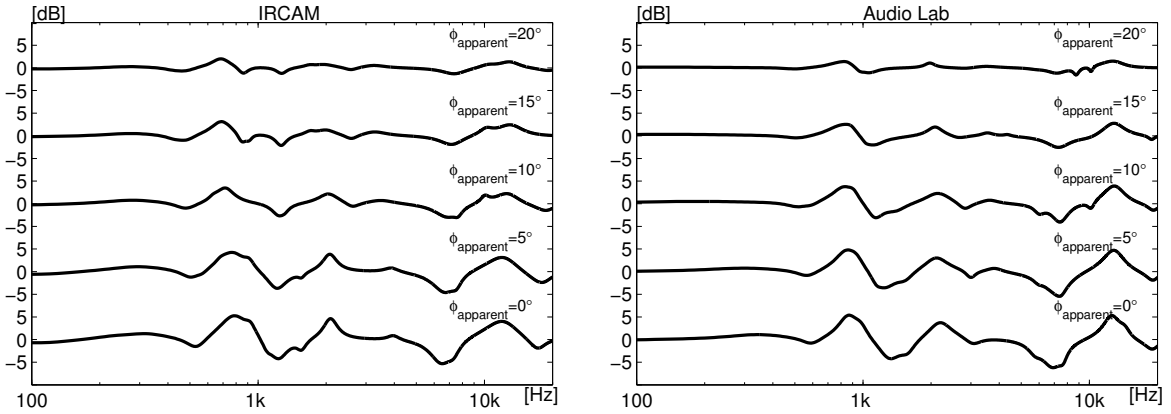


Fig. 8: Left column final IRCAM de-elevation filters. Right column final Samsung Audio Lab de-elevation filters. $\phi_{actual} = 25°$, curves are shifted by 15 dB for visualization.

or desired elevation angle ϕ of the sound source, and the denominator is the HRTF corresponding to the actual or physical location of the sound source. For all de-elevation filters, the azimuth angle θ is zero, corresponding to the frontal incidence direction. For each subject a de-elevation filter was created.

$$H_{de-el}\left(\phi_{apparent}, \phi_{actual}\right) = \frac{HRTF\left(\phi_{apparent}\right)}{HRTF\left(\phi_{actual}\right)} \quad (3)$$

The next step in creating a universal filter to perceive an apparent sound source de-elevated from its actual position is to average the de-elevation filters of all subjects on each database (IRCAM and Samsung Audio Lab). Left and right ear de-elevation filters for all subjects were grouped to obtain an average of all de-elevation filter magnitudes. In this paper, the filters were normalized for the case of a sound source placed at ϕ_{actual} = 25° elevation. Five de-elevation filters were computed for $\phi_{apparent}$ = 0°, 5°, 10°, 15°, and 20° (Fig. 7). As explained in Section 1, vertical localization relies most-ly in monoaural spectral cues, therefore it is valid to implement the de-elevation filter as a minimum phase approximation. An advantage of using an infinite impulse response (IIR) implementation is that one can modify the filter parametrically for different purposes. After obtaining an average across all subjects, the magnitude response of each de-elevation filter is characterized by a number of second-order sections (biquads) in cascade. The process for designing the PEQs was to first invert the magnitude response of the filter then setting a flat target at 0 dB from 20 Hz to 20kHz (Fig. 9). The constrained brute force (CBF) algorithm was then employed to minimize the error between target and

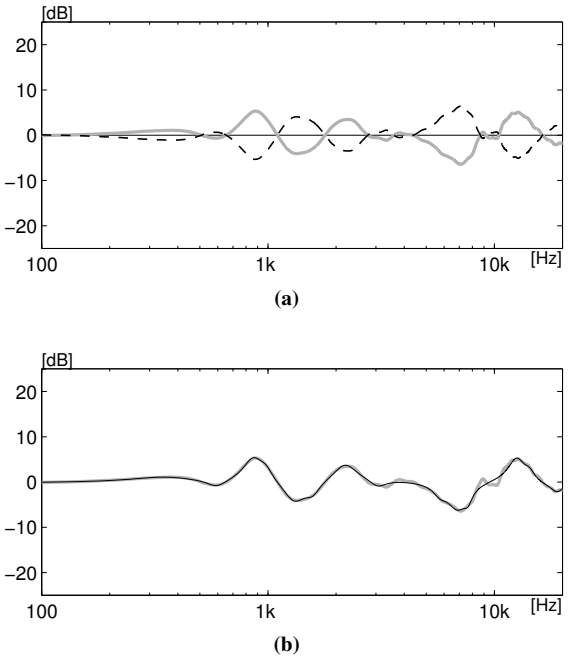


Fig. 9: De-elevation filter, Samsung Audio Lab data set. Plot (a) Gray line, original, black dashed line, inverted version of original. Plot (b) Gray line, original de-elevation filter, solid black line version approximation with 20 biquads.

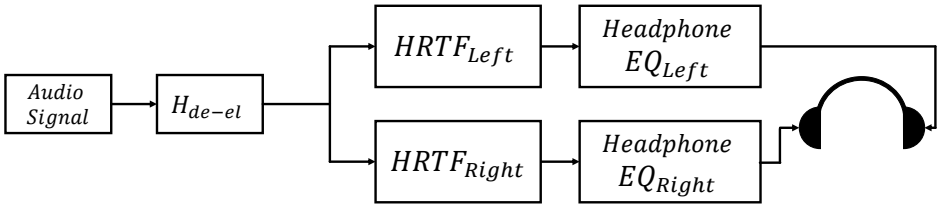


Fig. 10: Block diagram of playback chain for filter evaluation.

magnitude response [8]. In plot (b) of Fig. 9, an example of filter conversion from its original magnitude into 20 second-order sections in cascade is shown. The data corresponds to a de-elevation filter from Samsung Audio Lab data set to create an apparent source at 0° from a sound source physically placed at $\phi = 25°$ elevation.

2.7 Filter Evaluation

Preliminary evaluation indicated that the filter produced an apparent source at lower elevation angle than the actual angle. To verify the effectiveness in a number persons, listening tests were conducted. Since HRTF data of the 14 subjects was already acquired, auralizations were computed with the individual HRTFs to evaluate the de-elevation filter's performance. A series of listening tests were run across a panel of 12 listeners. Nine of the listeners were considered trained, based on performance in previous listening experiments, and all had been measured for normal audiometric hearing. The testing was administered using custom software made in Max 7®, to perform signal processing in real time (Fig.10). The software incorporated a video-tracked laser pointer to record the angle of elevation that the listener selected. This allowed for the listener to directly point onto a wall at the angle they perceived the sound to be emanating from. All tests were run over Beyerdynamic DT-990 Pro headphones equalized to each individual listener for proper binaural auralization per Møller et al. [9].

Prior to official testing, each listener underwent a familiarization session in which they moved the laser pointer vertically along a wall marked with elevations in 10 degree increments. As they listened to each of the three audio programs used in all the tests (pink noise, English male speech, and English female speech), the audio was filtered using the listener's individualized HRTF at 5-degree angles based on where he or she pointed on the wall. This allowed the listener to become accustomed to spectral changes in elevation through the test setup. Following familiarization, listeners took



Fig. 11: Left, blocked ear canal and miniature microphone. Right, headphone equalization process.

a test to evaluate their ability to vertically localize using their custom HRTF filters. Listeners compared their zero degree reference HRTF to six possible randomized HRTF elevation angles (0°-50° in 10° increments). The listeners completed an 18-trial test (3 programs × 6 angles) where both the program playback order and angle selected were randomized. In each trial, listeners pointed with a laser pointer to the angle of elevation from where they heard the audio. This angle was recorded by the software. Finally, listeners completed a two-session experiment to evaluate the performance of the de-elevation filters based of the IRCAM and Samsung Audio Lab databases. In these sessions, listeners compared their own 25°reference HRTF to a randomized treatment which could be either a de-elevation filter predicted to lower the audio to 0°, 10° or 20°, or the same 25° reference HRTF. In one of the test sessions, listeners heard the Samsung Audio Lab derived version of the de-elevation filter, while in the other session they heard the IRCAM database derived ver-

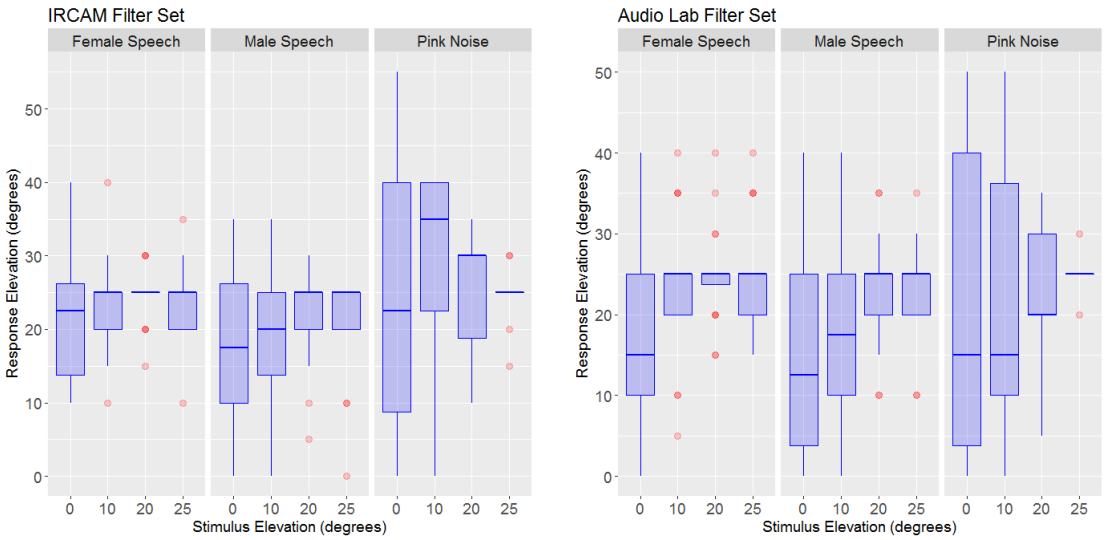


Fig. 12: Box plot results by track. Left, IRCAM de-elevation filter. Right, Audio Lab de-elevation filter.

sion. Each session consisted of 24 trials (3 programs $\times$ 4 treatments $\times$ 2 repeats) where the program order and de-elevation filter selected were randomized. Half of the listeners started on the test sessions, evaluating the IRCAM filters, while the other half started on the "Samsung Audio Lab" sessions.

3 Results

In Figures 12, and 13, box plot results of evaluation of the de-elevation filter by track, and by filter type, are shown. The boxes depict, interquartile range (middle represents 50% of responses), red-dots are outliers, and horizontal lines are the medians. In Fig.12 comparison of the de-elevation filter based on IRCAM, and Samsung Audio Lab databases is presented.

4 Discussion

In the tests validating individual HRTF filters, listeners tended to overestimate the elevation of pink noise tracks and underestimate the elevations of speech tracks. Variance was lowest on the male speech track (SD = 12.94) and highest on the female speech track (SD = 16.47). In the tests evaluating the IRCAM and Samsung Audio Lab based de-elevation filters, listeners generally heard both versions as de-elevating the sound. They tended to overestimate the elevation of the pink noise and

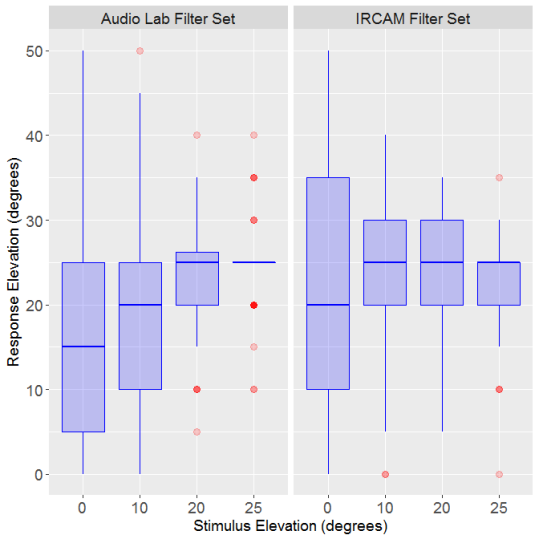


Fig. 13: Box plots results by de-elevation filter type.

variance was highest on this track (SD = 15.3). Interestingly, this was more of an issue with the IRCAM database version of the de-elevation filters. The Audio Lab de-elevation filter had a stronger tendency to de-elevate the signal across listeners than the IRCAM database version. On average, the Samsung Audio Lab filters tended to de-elevate sound 2 degrees more than the IRCAM database version, but on the pink noise track, the

difference was closer to 4 degrees more de-elevation. One can argue that this result was because, for the panel of subjects in test, the de-elevation filters were coming from average filters of the same data set so the probability of having a filter close to their own was higher.

This study was based on the physical characteristics of directional hearing included in the HRTF, but one has to remember than human sound perception in the median plane relies also on familiarity with the sounds and expectation. Another point to emphasize is that to perceive elevation or de-elevation, the sounds should contain wideband and dense spectrum. Due to time constraints, the evaluation test only contained male speech, female speech, and pink noise. Future work will include the evaluation of de-elevation filters with real loudspeakers in a typical scenario (e.g. a movie theater or living rooms). Another issue is the spectral change introduced by the filters. This was reported by the listeners and it was also detected in informal listening that the coloration introduced by the filters changed the sound quality to some extend. Optimization of the de-elevation filters can be done by detecting which part of the filter is strictly necessary to perceive the de-elevation effect and what part of the filter is not necessary. This would help to reduce unnecessary spectral coloration.

5 Conclusion

Measurements of HRTFs were performed in the frontal median plane on 14 subjects at the Samsung Audio Lab in Valencia, CA. Additionally the IRCAM database was included in the HRTF analysis to help understand human directional hearing in the median plane. Two sets of filters from the IRCAM and Samsung Audio Lab databases were constructed to create an apparent sound source de-elevated from its actual elevation angle of $\phi = 25^\circ$. The evaluation of de-elevation filters was carried out using binaural technique on the same measured subjects from Audio Lab data set. Results showed that a universal de-elevation filter for an actual source at 25°can create an apparent source de-elevated by about 13° for male speech and by 10° for female speech (right plot, Fig.12). The de-elevation filter from the Samsung Audio Lab database performed slightly better than the one from the IRCAM data set. As expected, the universal filter worked better for some subjects than for others due to anthropomorphic differences especially pinnae, head, and torso.

6 Acknowledgments

Samsung Electronics and Samsung Research America supported this work. The authors would like to thank the entire staff of Samsung's US Audio Lab for participating in the listening tests and Floyd Toole for his insightful suggestions.

References

[1] Møller, H., Sørensen, M. F., Hammershøi, D., and Jensen, C. B., "Head-Related Transfer Functions of Human Subjects," *J. Audio Eng. Soc*, 43(5), pp. 300–321, 1995.

[2] J. Blauert, *Spatial hearing – The psychophysics of human sound localization*, The MIT Press, 1999.

[3] Lopez, J. J., Cobos, M., and Pueo, B., "Influence of the Listening Position in the Perception of Elevated Sources in Wave-Field Synthesis," in *Audio Engineering Society Conference: 40th International Conference: Spatial Audio: Sense the Sound of Space*, 2010.

[4] IRCAM, "Listen HRTF Database," `http://recherche.ircam.fr/equipes/salles/listen/index.html`, 2002, [Online; accessed 19-July-2017].

[5] Lehtinen, A., "3D Modeling a Human Head," `http://www.secondpicture.com`, 2007, [Online; accessed 28-July-2017].

[6] Farina, A., "Simultaneous Measurement of Impulse Response and Distortion with a Swept-Sine Technique," in *Audio Engineering Society Convention 108*, 2000.

[7] Shaw, E. A. G., *The External Ear*, pp. 455–490, Springer Berlin Heidelberg, Berlin, Heidelberg, 1974, ISBN 978-3-642-65829-7, doi: 10.1007/978-3-642-65829-7_14.

[8] Ramos, G. and López, J. J., "Filter Design Method for Loudspeaker Equalization Based on IIR Parametric Filters," *J. Audio Eng. Soc*, 54(12), pp. 1162–1178, 2006.

[9] Møller, H., Hammershøi, D., Jensen, C. B., and Hundebøll, J. V., "Transfer Characteristics of Headphones: Measurements on 40 Human Subjects," in *Audio Engineering Society Convention 92*, 1992.