Audio Engineering Society

Conference Paper 20

Presented at the 6th International Conference on Audio for Games
2024 April 27–29, Tokyo, Japan

Significance of representing torso and individualized head location in HRTF

Jaan Johansson[1], Aki Mäkivirta[2], and Matti Malinen[1]

[1]*Kuava Oy, Kuopio, Finland*
[2]*Genelec Oy, Iisalmi, Finland*

Correspondence should be addressed to Aki Mäkivirta (aki.makivirta@genelec.com)

ABSTRACT

The significance of torso representation in a head-related transfer function (HRTF) is studied in this work. The Kemar HATS head position is modified in forward-backward and up-down directions according to the actual distribution of the human head location relative to the torso, and these are compared to the absence of torso to understand the audibility of the torso in the HRTF for different head positions. The spectral difference due to the absence of torso can exceed 1 dB in all sound arrival azimuth directions while the spectral difference decreases with increasing source elevation. The forward-backward head location has the strongest influence in the HRTF change. The importance of various areas on the torso reflecting sound is studied. A subjective listening test with personal HRTF demonstrates that the absence of torso is audible as sound colour and source location changes.

1 Introduction

Head related transfer function (HRTF) and head related impulse response (HRIR) describe the personal effect our bodies have to sound that is received at our ears. This effect is described in terms of binaural cues, chiefly inter-aural time difference (ITD) and interaural level difference (ILD), and monaural spectral cues. Binaural cues are significant in localizing sound sources on the horizontal plane [1] while spectral cues are utilized for sound localization in elevation [2]. The HRTF data intended for binauralization may present a torso usually fixed to the same direction with the head, or not present a torso at all. Also, binaural recordings may be created with simulators that represent only the head and no torso. The torso effect can become significant for HRTF modeling for headphone binauralization. Algazi et. al. [3] and Brown et al. [4] demonstrate the torso-related effect in the HRIR. The torso causes a second, delayed wave in the HRIR. In this work we consider the personal variation in the absence of torso in binaural rendering. We use the spectral difference between the situation representing a torso in the typical orientation, fixed in the same direction with the head, and we compare this to the absence of torso, i.e. with only the head being represented. The spectral differences between representing or not representing the torso are compared within the known range of the realistic head

locations relative to the torso. Also, the importance of various areas on the torso reflecting sound is studied by dividing the important sound reflecting area on the torso to smaller areas, and their contributions are investigated separately. The findings are supported by a subjective test on the audibility of the absence of a torso in a specific directions and using different types of audio signals.

2 Methods

2.1 Head Position

The definitions for the head and torso locations are demonstrated in Fig. 1. Using the photogrammetric method described in [5], fixed-torso 3D models of 195 actual persons are acquired and then their head locations are determined [6] (Fig. 2). The head location is taken as the midpoint between ear canal entries (black dot), also taken as the origin of the coordinate system, $(x_H, y_H, z_H) = (0,0,0)$. The torso location relative to the head is then detected using the methods detailed in [6]. The torso location (black triangle) is determined by finding the highest point (crest) over the shoulders and their joining line and then determining the x, z value at $y = 0$.

2.2 No Torso and Modified Torso Locations

A fixed-torso 3D model of the Kemar HATS is acquired using photogrammetry [5] having torso in the same direction with the head. The head-only model of the Kemar HATS is obtained by separating and removing the torso at the neck in the 3D model. The translated torso versions of the Kemar are obtained by translating the torso in x and z directions and rejoining the torso and head with a naturally shaped smooth neck section into a new 3D model.

2.3 Spectral Difference

HRTFs for the required source azimuth-elevation positions are obtained using boundary element method (BEM) [7, 8] applying sound-hard boundary condition. Using the sound-hard surface impedance may somewhat emphasize the spectral features over using a surface impedance more similar to skin and textile. However, in this work the sound-hard acoustic impedance removes the question about the values that should be used to best model skin, hair or clothing on a person.

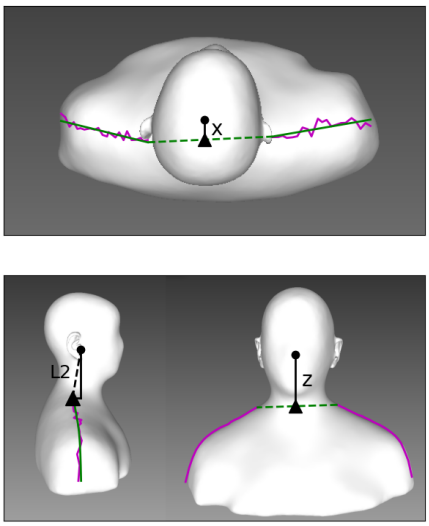

Fig. 1: Determining the head position in the 3D model. Head position (black dot) is the midpoint between ear channel entries and the origo of the coordinate system. Raw shoulder crest locations (purple). The shoulder crest interpolation across the neck area (green). Torso location (marked by triangle) is described by the x, z coordinates (black solid) and *L2* (black dashed). The example shows the Kemar HATS. Figure adopted from [6].

The spectral difference S is obtained by comparing the head-only HRTF H_h to the HRTF with torso $H_{o(x,z)}$,

$$S_{x,z}(f,\phi,\theta) = \frac{H_h(f,\phi,\theta)}{H_{o(x,z)}(f,\phi,\theta)}. \quad (1)$$

The magnitude and phase of the spectral difference S is studied for complete 360 deg azimuth range ϕ and source elevations $\theta = [-20...+60]$ deg. The significance of the head position to the torso effect is studied by analysing the spectral difference using modified Kemar HATS 3D models. These models are generated by translating the torso in x and z directions. The torso translations are chosen such that the head position range found in the studied population is covered.

The spectral difference is studied in the frequency range $f = [50...6000]$ Hz which includes and exceeds the frequency range previously appointed to the torso [9, 10, 11] while avoiding excessive computational workload. It is generally accepted that local increases

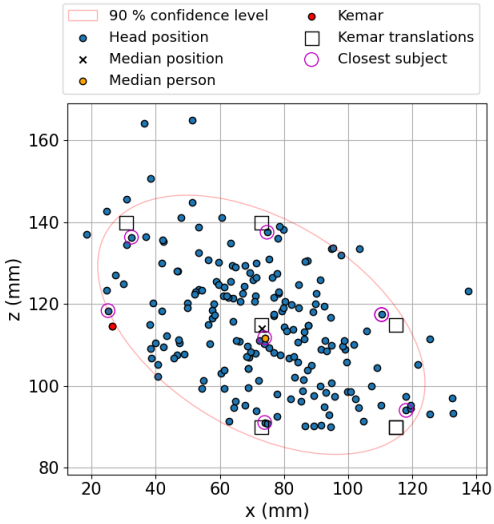

Fig. 2: Personal head positions relative to the shoulders for N=195 persons (blue dots). The 90% confidence limit for the bivariate normal distribution representing this data is depicted with the red ellipse. The head position of original Kemar HATS (red dot), and translated Kemar HATS cases (black squares). Figure adopted from [6].

and decreases of level can become detectable when their magnitude exceeds about 1 dB, with a slightly smaller detectability for the notches or local decreases in level [12, 13]. We plot the changes in level larger than 1 dB to indicate the changes in levels of the spectral difference that can become audible. The limit of 10 degrees for the phase has been found to represent the limit of audibility [14, 15] and this limit is used in the same way, with phase changes larger than this limit being presented. These two limits mainly aid the visual inspection of magnitude and phase difference plots.

2.4 Significance of torso areas

Once the total contribution of the torso to the HRTF is described using the spectral difference, the role of the different parts of the torso are studied. To study the torso's effect further, 17 areas of the Kemar torso are selected (Fig. 3) and their effect to the total wave field, measured as the HRTF, are examined. The measured total wave field at the ear channel entry consists of the initial wave field and the scattered wave field. To understand the contribution of a patch, after calculating the total pressure on the 3D model, the pressure in the patch area is set to zero, and the pressure at the ear channel entry produced by the modified total pressure field is computed, removing the scattered wave field contributed by this patch area.

Fig. 3: 3D model of Kemar with 17 separate patches colored.

2.5 Subjective Listening Test

A blind pair-wise A/B listening test was conducted to determine the audibility of the absence of torso. The cases tested were the torso aligned with the head (later called 'normal'), and the head alone, without a torso (later called 'head'). Each of the persons participating were individually modelled for personal HRTF by creating a 3D model of the person's head with upper torso aligned with the head using photogrammetric capture of shape ('normal' case). This 3D model was then modified for the no-torso ('head') case. The personal HRTFs were then calculated at 0, 30 and 60 deg azimuth at 0 deg elevation by modelling the sound field using the sound-hard acoustic impedance assumption for the head and torso, leading to slightly enlarged diffraction effects somewhat accentuating spectral cues. The azimuth direction 0 deg is directly in front of listener, and 30 and 60 degrees move the sound source to the left of the listener. The listening test material was prepared using these personal HRTFs, resulting in binaural stereo tracks representing the samples for 'normal' and 'head' cases in these three directions.

The audio material consists of ten-second audio samples of music and male speech, originating in the EBU SQAM disk [16], as well as pink noise. The audio samples were adjusted for similar subjective loudness,

enabling listening without adjusting the output level throughout the listening test.

The material was delivered to each participant as a DAW project file containing 16 unique binauralized track pairs {normal,normal} and {normal,head}, for all the audio signal type and direction of the sound source combinations. Each binauralized track pair was identified by a sequence number, but the order of presentation was randomised individually for each listener. The participants were encouraged to listen through the test material to learn about the nature of the possible differences audible in track pairs. The number of repetitions of each track pair during evaluation was not limited. Repeated listening of the track pairs was encouraged to increase the ability to hear even small differences.

The listener entered the evaluation of the track pair difference using a discrete scale from 0 to 3, defined as follows 'no audible difference' (value 0), 'a small audible difference' (1), 'clear audible difference' (2) and 'a very significant audible difference' (value 3). Only integer values 0 to 3 were given in the responses. The responses were collected from listeners using a spreadsheet file, showing the track pair number and evaluation. The mean of the difference score for the normal and missing torso cases were calculated for the audio types and directions of audio presentation. Non-parametric hypothesis tests were performed to detect the significance of the differences seen in the means. The level of significance is generally taken to be 0.05 but the hypothesis risk levels are separately reported for each test. Non-parametric hypothesis testing is used because of the sparsity of the grading scale and the limited size of the test panel.

3 Results

3.1 Spectral Difference

The spectral difference studies how the presence of the torso changes the HRTF at the various head positions. The six modified Kemar models cover the realistic human head position distribution in the horizontal position (x coordinate) and vertical position (z coordinate). The unmodified Kemar HATS head position is very upright, placing the head almost on top of the torso, at very low x coordinate values. The torso acts to reflect and diffract audio energy towards the ears. This causes a time delay dependent on the distance of the head relative to the torso. The effective area of the sound reflecting surface on the torso affects the level of the reflected or diffracted sound. The reflected or diffracted sound creates a spectral change similar to that of a comb filter. The spectral difference elicits this filter, and this filter has both magnitude and phase characteristics of the change between having the torso and head or only having the head.

The magnitude of these spectral differences are shown in Fig. 4. The forward head positions relative to the torso cause the peaks and valleys in the spectral difference magnitude to move towards lower frequencies while at the same time the peak-and-valley structure becomes more dense in frequency. This gives rise to a higher number of peaks and valleys in the studied frequency range when the head moves forward relative to the torso. A similar feature is also seen in the spectral difference also when the head moves higher relative to the torso.

As the HRTFs are considered to be minimum-phase systems [17] the spectral difference displays the minimum-phase characteristics, also. The phase of such a system is linked to the magnitude. The head position relative to the torso will produce effects in the phase of the spectral difference linked to its magnitude. This is evident in the phase of the spectral difference (Fig. 5) where typically the maximum deviations in the phase are close to the frequencies where the magnitude of the spectral difference crosses the unity gain.

The spectral difference was evaluated for certain values of elevation to demonstrate the nature of change in the spectral difference with changing elevation (Fig. 6). The increasing elevation reduces the manifestation of the spectral difference particularly towards higher frequencies and on the side opposite to the sound source while at lower frequencies the reduction is small and there is even minor local increase below 1 kHz particularly for the model describing the mean human head position relative to the torso.

Angle-spectrum area (Fig. 7) describes how widely the significant spectral difference exists across the studied frequencies. It integrates the area constituted by the azimuth angle and spectrum where the spectral difference exceeds 1 dB. This is a relative measure where 100% would mean that a spectral difference exceeding 1 dB would be seen in all azimuth angles and across the complete range of the studied frequencies. The angle-spectrum value decreases with the increasing elevation, indicating that the spectral difference is increasingly

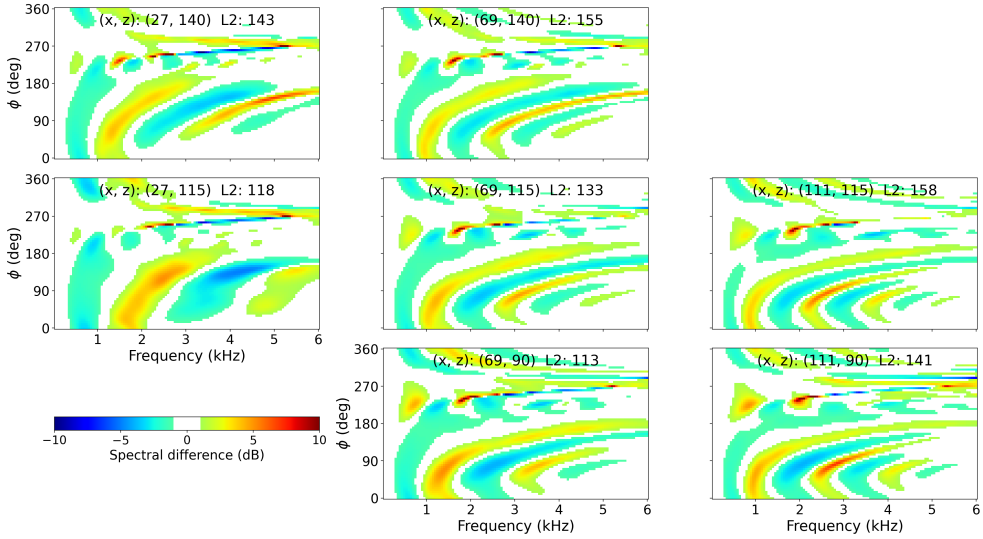

Fig. 4: The magnitude of the left ear spectral difference between full Kemar and Kemar without a torso. The standard Kemar head location is shown on middle line, left panel. The panels in the center and right columns show head positions moved forward relative to the torso. The top row shows head positions moved up. The bottom row shows the head positions moved down. Spectral difference magnitude less than 1 dB is not shown.

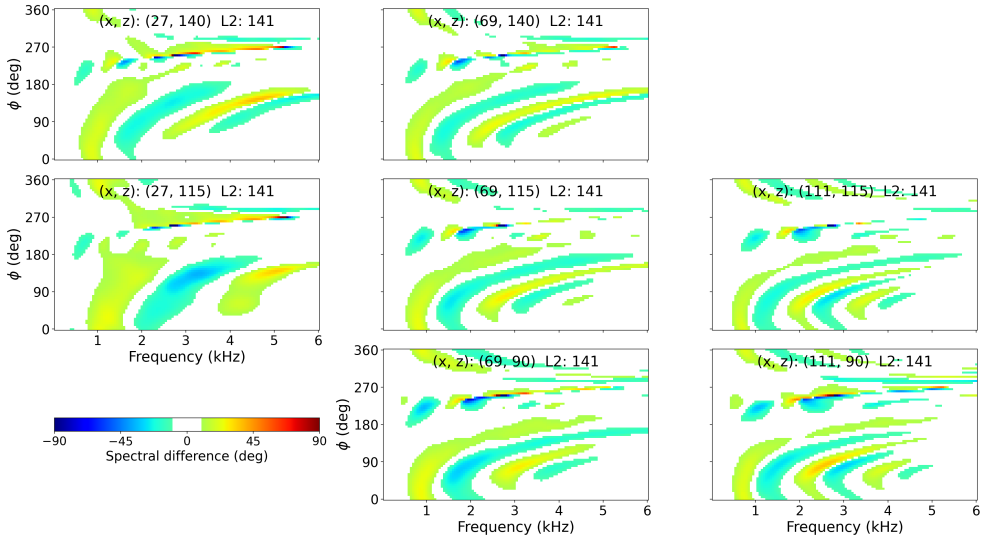

Fig. 5: The phase of the left ear spectral difference presented similarly as in Fig. 4. The spectral difference angle less than 10 deg is not shown.

falling below 1 dB or covering less frequencies or the azimuth angle range. The spectral changes reduce overall with increasing elevation, implying that the significance of the torso is smaller at higher elevations.

3.2 Significance of torso areas

The patch torso model represents a full torso model where the contribution of the surface pressures in the

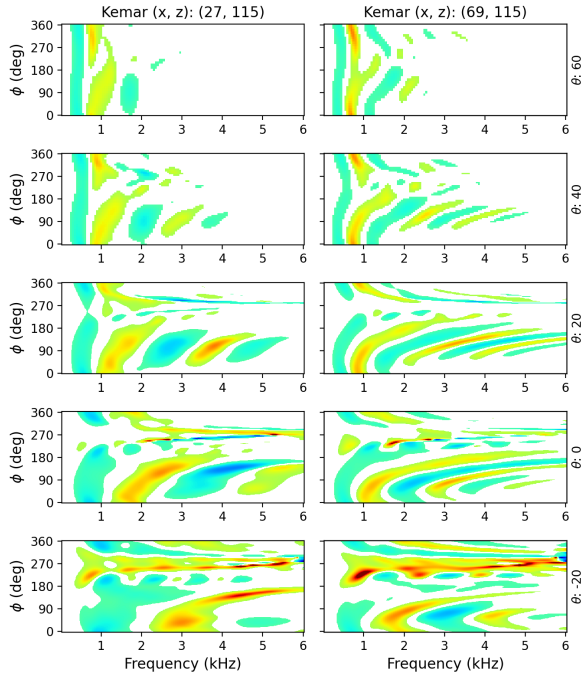

Fig. 6: The effect of elevation θ to the magnitude of the spectral difference across the azimuth angles ϕ. The unmodified Kemar HATS (left column) and Kemar HATS head moved to the typical human head position (right column). The elevation θ ranges from +60 deg (top) to -20 deg (bottom), indicated on the right side of the panels.

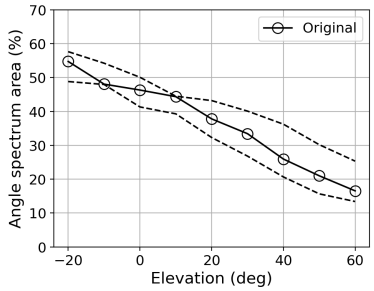

Fig. 7: Angle-spectrum area exceeding 1 dB as a function of sound source elevation. Unmodified Kemar HATS (solid) and the maximum and minimum (dashed) of the unmodified as well as the six modified head locations taken at the normal distribution 90% confidence limit. See also Fig. 2.

combined colored area (the patches, Fig. 3) have been removed in the HRTF calculation. The spectral difference between the patch torso model and full torso model is depicted in Fig. 8. Comparing this to the middle row left panel in Fig. 4, showing the effect of full removal of the torso, it is evident that the patch areas contribute most of the torso-related effect seen in the HRTF. The significance of the various areas in the torso (the patches) varies with frequency and direction of arrival of the audio signal. The relative effect of each of the patches is shown in Fig. 9, with the back side of the shoulder (green colours) contributing the least and the front top of the chest (red colours) having the largest contribution, and the areas closer to the ear having larger relative contribution.

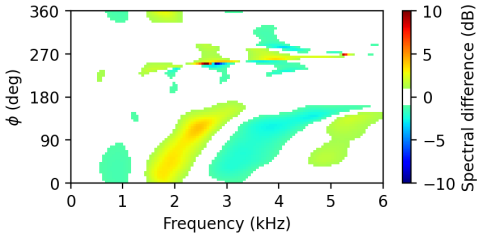

Fig. 8: The magnitude of the spectral difference between the HRTF with the contribution of all the patches excluded (Fig. 3) and the full Kemar model HRTF.

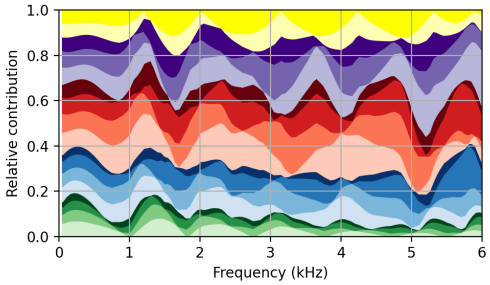

Fig. 9: The relative contribution of the surface pressure of the patch areas in Fig. 3 to the Kemar left ear HRTF for audio arriving from directly ahead $(\phi, \theta) = (0, 0)$.

3.3 Subjective Listening Test

Eight male participants (28...63 yrs) judged how similar a pair of audio tracks is for 27 unique pairings

comparing 'head' (missing torso) to the 'normal' case (head with torso), using over-the-ear headphones in a quiet room. Each participant adjusted the sound level to a comfortable level allowing access to all details in audio although the sound level was not verified by measurement. The audio interfaces used were Focusrite Scarlet Solo (1), Scarlett 2i2 (2), HDA-Intel HD-Audio (generic) (3), MOTU Microbook ASIO (4), Realtek ALC892 (5), Apollo Twin X (6). The headphone types used included AKG 240 (5), HI-X55 (6), DT 770 Pro (80 Ohm) (2), Sennheiser HD 555 (5) and HD 600 (1,4), WH-1000XM2 (used with cable) (5), and Superlux HD-681 (3). The numbers in parenthesis identify match between headphones and audio interfaces.

	Normal-Normal	Normal-Head
Noise	0.13±0.34	2.46±0.78
Speech	0.25±0.53	0.96±0.86
Music	0.25±0.44	1.58±0.93
	Normal-Normal	Normal-Head
0 deg	0.08±0.34	1.75±1.19
30 deg	0.25±0.53	1.33±0.92
60 deg	0.29±0.46	1.92±0.97

Table 1: Difference scores, mean± SD

The mean difference scores with standard deviations (SD) for the audio type and direction of audio presentation are summarized in Table 1 for the head with torso case compared to itself ('Normal-Normal') and to the head without torso case ('Normal-Head'). For the audio type, all directions of presentation are pooled. For the direction of presentation, all audio types are pooled. The 'Normal-Normal' comparison served as an anchor case to indicate how sensitive our listening panel was in detecting the presentation of a test sample pair with no difference, mostly correctly identified (grade 0), or as 'small difference' (grade 1). In the direction 0 deg (in front of the listener) the 'Normal-Normal' case was detected the best (mean difference score 0.08) with the mean difference score increasing with angle (30 deg, score 0.25 and 60 deg score 0.29). Using the Kruskal-Wallis test on these angles followed by pairwise Dunn-Bonferroni test does not detect significant differences of these mean values. This demonstrates that the 'Normal-Normal' case was correctly detected.

The lack of torso was audible for all tested audio signal types and in all tested directions of audio presentation. The 'Normal-Head' comparison for the 'speech' sample was the hardest to judge (mean difference score 0.96) and the 'noise' sample produced the clearest differences (2.46). For all the test signals, the non-parametric Mann-Whitney U test finds significant difference, for 'noise' and 'music' at risk level $p < 0.001$, and for 'speech' at $p = 0.004$. Lack of torso was audible for all the tested directions. 'Normal-Head' cases in directions 0, 30 and 60 deg are significantly different at risk level $p < 0.001$ using the non-parametric Mann-Whitney U test. The nature of the changes in audio reported by the listeners were predominantly spectral differences, i.e. changes in sound colour, although also positional changes were noted.

4 Discussion

The torso-related spectral difference has complex characteristics that are linked to the position of the head relative to the torso. The forward-backward head position has the largest impact to the acoustic imprint of the torso in the HRTF. While Gardner et. al. [11] find the torso dominates localization below 2 kHz frequency, the spectral difference is seen to extend across the frequency range to high frequencies but is not significant below about 300 Hz. The spectral difference is linked to a potential change in sound color. The spectral difference also shows phase-related effects that are closely linked to the magnitude of the spectral difference as the HRTF is closely a minimum phase system response [17]. The maximum phase difference typically coincides with zero crossings in the spectral difference magnitude.

Evaluation of the importance of the different areas of the torso indicates that the torso area in the direction of sound arrival are significant to the torso effect.

The subjective listening test indicates that the torso appears audible in the all the tested directions of audio and for all the tested types of audio (pink noise, music, speech), and is described as sound colour changes while location changes where also reported, similar to Brinkmann et. al. [18]. The spectral difference magnitude decreases with increasing sound source elevation having the maximum at the lowest elevation angle studied.

The use of sound-hard acoustic impedance may slightly exaggerate the spectral difference, which may be slightly reduced if the real acoustic impedances are used for the skin, hair, and torso with clothing. However, the difference between these and using the sound-hard impedance does not appear very significant [19].

As the torso is audible in the HRTF in the directions and with the audio signal types studied, it appears that use of head simulators without a torso and HRTFs taken out of such simulators should be discouraged.

5 Summary

The torso is audible in HRTF for the studied directions. The spectral difference due to the torso is mainly influenced by the forward-backward location of the head relative to the torso, while up-down head location has less effect.

References

[1] Macpherson, E. A. and Middlebrooks, J. C., "Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited," *J. Acoust. Soc. Am.*, 111(5 Pt 1), 2002.

[2] Watkins, A. J., "Psychoacoustical aspects of synthesized vertical locale cues," *J. Acoust. Soc. Am.*, 63(4), 1978.

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

[4] Brown, C. P. and Duda, R. O., "A Structural Model for Binaural Sound Synthesis," *IEEE Transactions on speech and audio processing*, 6(5), 1998.

[5] Mäkivirta, A., Malinen, M., Johansson, J., Saari, V., Karjalainen, A., and Vosough, P., "Accuracy of photogrammetric extraction of the head and torso shape for personal acoustic HRTF modeling," in *148 AES Conv.*, 2020.

[6] Johansson, J., Mäkivirta, A., and Malinen, M., "Torso Effects in HRTF," in *Audio Engineering Society Conference: AES 2023 International Conference on Spatial and Immersive Audio*, 2023.

[7] Johansson, J., Mäkivirta, A., Malinen, M., and Saari, V., "ITD Prediction Using Anthropometric Interaural Distance," *J. Audio Eng. Soc.*, 70(10), 2022.

[8] Young, K. and Kearney, G., "A High-Resolution Boundary Element Method Suitable Full Torso Mesh of KEMAR," *J. Aud. Eng. Soc.*, 71, 2023.

[9] Algazi, V. R., Duda, R. O., Duraiswami, R., Gumerov, N. A., and Tang, Z., "Approximating the head-related transfer function using simple geometric models of the head and torso," *J. Acoust. Soc. Am.*, 112(5), 2002.

[10] Kuhn, G. F., "Physical acoustics and measurements pertaining to directional hearing," *105th Meeting: Acoust. Soc. Am.*, 73, 1983.

[11] Gardner, M. B., "Some monaural and binaural facets of median plane localization," *J. Acoust. Soc. Am.*, 54(6), 2015.

[12] Moore, B., Oldfield, S., and Dooley, G., "Detection and Discrimination of Spectral Peaks and Notches at 1 and 8 kHz," *J. Acoust. Soc. Am.*, 85(2), 1989.

[13] Toole, F. and Olive, S., "The Modification of Timbre by Resonances: Perception and Measurement," *J. Aud. Eng. Soc.*, 36(3), 1988.

[14] Hansen, V. and Madsen, E. R., "On Aural Phase Detection: Part II," *Journal of Audio Engineering Society*, 22(10), 1974.

[15] Cabot, R., Dorans, D., Mino, M., Tackel, I., and Breed, H., "Detection of Phase Shifts in Harmonically Related Tones," *Journal of Audio Engineering Society*, 24(7), 1976.

[16] The European Broadcasting Union, "EBU SQAM CD," https://tech.ebu.ch/publications/sqamcd, accessed Oct. 26, 2023.

[17] Kulkarni, A., Isabelle, S., and Colburn, H. S., "On the Minimum-Phase Approximation of Head-Related Transfer Functions," in *Proc. 1995 Workshop on Applications of Signal Processing to Audio and Accoustics*, 1995.

[18] Brinkmann, F., Roden, R., Lindau, A., and Weinzierl, S., "Audibility and Interpolation of Head-Above-Torso Orientation in Binaural Technology," *IEEE Journal Of Selected Topics In Signal Processing*, 9(5), 2015.

[19] Katz, B., "Boundary element method calculation of individual head-related transfer function. II. Impedance effects and comparisons to real measurements," *Proc. Workshop on Applications of Signal Processing to Audio and Acoustics*, 110, 2001.