TOWARDS A HUMAN PERCEPTUAL MODEL FOR 3D SOUND LOCALIZATION

S. Mattes Institute of Sound and Vibration Research, University of Southampton, UK
P.A. Nelson Institute of Sound and Vibration Research, University of Southampton, UK
F.M. Fazi Institute of Sound and Vibration Research, University of Southampton, UK
M. Capp Research Department, Meridian Audio Ltd, UK

1 INTRODUCTION

This paper describes the theoretical fundamentals and the preliminary results of the implementation of a perceptual model for human sound localization for the evaluation of 3D audio systems. Recent work undertaken in ISVR[33] has led to the development of a biologically inspired auditory signal processing model that gives remarkable accuracy in the prediction of human localization of acoustic sources. Nevertheless, the model relies only on some auditory cues, i.e. the binaural cues, and as a consequence is able to predict only the location of sources in the horizontal plane. This paper reviews the basis of the previous model and presents preliminary results of an enhanced perceptual model that takes into consideration other cues, such as the spectral and dynamic cues, in addition to the classical binaural cues, in order to predict the location of 3D sounds.

2 OVERVIEW

2.1 Sound Localization

Sound Localization is a perceptual process, law or rule, by which the location of an auditory event (defined by its perceived position relative to a listener) is related to specific attributes of a sound event, or of another event that is in some way correlated with the auditory event[10]. In contrast to other sensory systems, like vision and taste, in sound localization there is no point-to-point correspondence between a sound event and the perceived locus (auditory event) of an acoustic image at the lower peripheral stages of the human hearing system. Instead, it is believed that the localization of sound events occur entirely as a consequence of neural processing of monaural or binaural signals[4].

Several types of information are involved in the perception of such an event, including visual and cognitive information in addition to auditory information. We divide the cues that are used for localization into three main categories: the static, dynamic and motion cues. Static cues include the well-known binaural cues associated with ITDs (interaural time difference) or IPDs (interaural phase difference) and ILDs (interaural level difference), the monaural spectral cues (also known as pinna cues due to the significant contribution of the pinna to the creation of spectral changes) and the distance (or intensity) cues. The dynamic and motion cues are the same as the static cues, with the difference that for motion cues all static cues change by changing the position of the sources, while dynamic cues are produced by changing the position of the listener.

2.2 Perceptual Models

We define a "Perceptual Model" as a simplified representation of a living organism, which reveals a similar behaviour to its prototype in the way information is interpreted and organized. In the case of a perceptual model for human sound localization, the main goal of the model is to be able to localize the source in a similar way to human beings.

Perceptual models can be divided into two main categories: statistical and biologically inspired.

In statistical (probabilistic or analytical) models, prediction relies heavily on stimulus statistics rather than on explicit waveforms. This class of model provides a powerful means to understand many aspects of data in the literature, independent of the details of realization. At the same time however, the analytical nature presents us with the limitation that it is very difficult to obtain predictions for arbitrary stimulus configurations[37]. Moreover, most of the statistical models act as perfect listeners[5, 42], while in other cases the model either overestimates or underestimates the location of the stimulus[31].

Biologically inspired models are based on neuroscientific findings and the elements that are used for their implementation are found in the neurophysiological pathways of hearing. One of the most common techniques that has been used in recent decades is based on interaural cross correlation, which is supported by the so-called EE (excitation-excitation) units. The majority of these models rely on Jeffress's coincidence detector[20] and its physiological basis are the so called EE-type cells which can be found in the MSO (medial superior olive) [8, 9, 22, 23, 44] of the central auditory system in the brainstem.

Some other models are based on (Durlach's) EC (equalization-cancelation) theory[16] which is supported by so-called EI (excitation-inhibition) units. The internal representation of EI cell activation is analyzed by a template-matching procedure[13, 33]. Although the EC theory is purely analytical, there is some physiological basis of EC-like processes in the mammalian auditory system, in the LSO (lateral superior olive) and a subgroup of cells in the IC (inferior colliculus) [8, 9, 12, 15, 36]. The main idea of this theory is that the cells are excited by the signals from one ear and inhibited by the signals from the other ear.

In recent years, several other models have appeared that have a similar structural representation of the human auditory pathway[34]. These models are based on the "count comparison principle" which provides an alternative to cross-correlation and is supported by the most recent neurophysiological findings.

2.3 Current Model

The model[33] used in this paper is based on EC theory for the production of the excitation-inhibition (EI) pattern in binaural processing. In particular, the model consists of three main stages, each of which corresponds to different (and more or less known) operations of the human auditory system in spatial hearing.

The model starts with the peripheral processor, which takes binaural signals as an input. This stage consists of a unit which corresponds to a time-invariant band pass filter from 1kHz - 4kHz with a roll-off of 6dB/octave below 1kHz and -6dB/octave above 4kHz, which represents the response of the human middle ear. This is followed by a fourth-order gammatone filterbank with 60 channels between 300Hz and 12kHz, which represent the frequency selectivity of the basilar membrane. Each gammatone filter output is processed by a half-wave rectifier, a fifth-order low pass filter with a cut-off frequency at 770Hz, and a square root compressor, which respectively represents the organ of Corti, the gradual loss of the phase-locking in neural firing, and the nonlinearities of the basilar membrane.

The model continues with the pre-processor, which is a binaural processor based on the EC theory for the extraction of the excitation-inhibition (EI) cell activity patterns (EI-patterns). Signals from one ear are compared to the corresponding signals from the other by means of EI interactions as a function of the internal characteristic ILD and ITD. This produces a time-dependent internal representation of the binaurally presented stimuli.

In the final stage is the central processor, which considers the EI cell activity pattern for the estimation of source location in the horizontal plane. This is a decision making device using a simple pattern matching process. This means that in order for the central processor to find the unknown location of the sound source it has to compare the given EI-pattern with a bank of EI-pattern templates for the horizontal plane.

3 TOWARDS A MODEL FOR ELEVATION

3.1 Choosing an appropriate coordinate system

The position of the sound event compared to the auditory event is usually made in terms of a head-related coordinate system (Figure 1, 2), i.e. the system of coordinates shifts in conjunction with movements of the subjects' head[10]. For this reason, the perception of sound localization can be divided into three perceptual sub-processes[26], which correspond to the three planes defined by the spherical coordinate system. The horizontal plane (or sagittal plane) perception, where the ITD and the ILD cues are considered the main cues for this perceptual process[10, 26, 30], the vertical plane (or median plane) perception, where the spectral cues are the main contributors of this perceptual process[19, 26] and the distance perception, where the intensity cues are considered most important[26].

Although the vertical-polar coordinate system (Figure 1) is the most well-known spherical coordinate system in audio engineering, the interaural-polar coordinate system (Figure 2) has been considered in the current paper to be more suitable for human perception of localization due to the nature of the interaural cues that are being used in horizontal plane perception. One of the most well-known phenomenon in spatial hearing is the inability of the human to localize a sound source at all the points that are at the same distance from both ears[10]. The locus of all these points form a cone, which is usually known as the cone of confusion[10]. This means that for sources lying on the a circle describing the base of the cone, the interaural differences (ITDs, ILDs) remain largely the same[10, 27]. This can be described better in the interaural-polar coordinate system since, a constant azimuth angles θ', describes the base of the cone of confusion (Figure 2).

However, the two spherical systems of (Figure 1, 2) are equivalent and can be transformed to one another by the equations

$$\sin\theta' = \cos\theta \cdot \sin\varphi \quad (3\text{-}1) \qquad \sin\theta = \cos\theta' \cdot \sin\varphi' \quad (3\text{-}2)$$

$$\cot\varphi' = \cot\theta \cdot \cos\varphi \quad (3\text{-}3) \qquad \cot\varphi = \cot\theta' \cdot \cos\varphi' \quad (3\text{-}4)$$

Where θ', φ' are the azimuth and elevation angle corresponding to the interaural-polar coordinate system and θ, φ are the elevation and azimuth angle corresponding to the vertical-polar coordinate system.

3.2 Preprocessing stage

The localization ambiguity arising from the cone of confusion can be resolved quite readily by head motion[10]. However, even if the head is restrained, partial resolution is still possible on the basis of the static spectral cues[6, 17, 18, 41]. Resolution of the ambiguity is further improved if the listener has *a priori* information which restricts the possible source locations. For example, if the subject knows in advance that the sound source is in the frontal horizontal plane.

Considering these factors, it was necessary to verify the ability of the current model to resolve any static cues of elevation. It has been found out from our data that the EI pattern cannot resolve any elevation cues and as a consequence it is necessary to further explore the cues that could give any indication of elevated sources. Analysis of HRTFs has shown that specific bandwidths are responsible for specific aspects of perception. For instance, the front-back locations, which cannot be resolved with just the ITDs and ILDs due to their ambiguation on a cone of confusion, can be discriminated with the appropriate spectral cues that reside mainly at 8-16kH [24, 28], while for up-down location the appropriate spectral cues reside mainly at 6-12kHz[24, 28].

For this reason, it was considered necessary to analyse known HRTF databases which include a wide range of different subjects and resolutions. In the current paper three different HRTF databases have been analyzed[1, 32, 40]. In addition to the CIPIC database[1] already expressed in the interaural-polar coordinate system, both the ISVR[32] and IRCAM/AKG[40] databases have also been transformed to the interaural-polar coordinate system.

3.3 Analyzing the HRTFs

The HRTFs have been analyzed in the literature both in the frequency, and in the time domain using HRIRs (head related impulse responses) (Figure 3 - 6). One of the most salient features on the analysis of HRIRs is the identification of the ITDs and ILDs, where the acoustic wave that first reaches the ipsilateral pinna appears to have higher magnitude and time advance from the corresponding acoustic wave on the contralateral ear (Figure 4). The ILDs can also been seen in the corresponding HRTFs (Figure 5) mainly at higher frequencies, above around 4kHz.

The spectral cues, shown in the HRTFs, can be divided into two main cues the head/torso and pinna cues. The pinna cues contribute to both spectral peaks and notches[35] and the notches can be found above 6kHz due primarily to the scattering of the acoustic wave by the pinna[39]. Although, the pinna cues are the most prominent cues for elevation, notches that are caused by head diffractions[11, 35] and torso reflections (1-3kHz), could be used as potential cues for vertical localization for low frequency sounds[2, 7] (such as footsteps or thunder that have low energy at high-frequencies[3]). More specifically, computed HRTFs[3] showed that the arch-shaped notches that are symmetric about 90° elevation angle are due to specular reflections from the upper torso. These are shaped as comb-filter notches through the spectrum (Figure 3b, 5) and appear at around 1.2ms in the time domain in the close vicinity of the direct wave (Figure 3a, 4). Moreover, the deeper notches around 210 to 250° are caused by torso shadow while at around 255° a torso bright spot appears. These low-elevation notches combined with head shadow, cause the response for frequencies above 1 kHz to be much lower on the contralateral side than on the ipsilateral side[3,39] (Figure 3b, Figure 5).

In the HRTFs (shown in Figure 3, Figure 5) peaks and notches can be found. However, due to the fact that the notch frequency varies smoothly with elevation, even for rather narrow notches[29], it is thought that this provides the main cue for the perception of elevation[35]. Notches are caused by multiple reflections from different parts e.g. head, torso, thighs (in case of seated listeners, indicated also as knee reflections in references [35, 39]) and pinna cavities. On the contrary, spectral peaks, or pinna resonances, do not show this smooth trend. However, it is likely that the presence or absence of the spectral peak could itself be a strong cue for elevation[35]. Peaks can been seen on the figures as bright patches[38].

A visual observation on the spectral notches of Figure 3, 5 reveals that their frequency changes with elevation. For instance, a pinna notch appears at 6kHz to 12kHz that moves with elevation at 45° azimuth angle[7] while in the median plane it is in the region of 6-10kHz (Figure 3b). These features are difficult to see in the time domain and are prominent notches above 5 kHz. For elevation from -45° to +90° there are three prominent notches attributed to the pinna (Figure 3). In these cases as the elevation increases the frequency of the notches also increases[35]. Note that the notch that appears just above 14kHz has been attributed to the probe that has been inserted at the ear canal for measuring the HRTF that results in a standing wave[14].

The pinna notches in the contralateral side (Figure 5b) can be explained if we assume that the sound diffracts around the head entering the contralateral concha at approximately the same elevation angle as if the source were in the ipsilateral hemisphere[25]. However, since elevation perception is essentially thought to be monaural[21] it is likely that humans use only the near ear (i.e., the ear closest to the source) for vertical localization. Moreover, although it is still possible that the pinna notches in the contralateral HRTF could provide extra cues for vertical localization[35], we can notice from Figure 5 that the difference in magnitude between the ipsilateral and contralateral ear is about 10-30dB for frequencies above 3kHz to 4kHz[43], which means that the ipsilateral ear has higher energy.

Finally, it is worth mentioning that the variations with elevation are different for the left and the right pinnae[35, 43] which can be explained by the fact that the left and right pinnae do not have the same shape and dimensions[1]. To what extent this variation influences the localization is not known.

4 DISCUSSION

It is well known that HRTFs play a central role in spatial audio systems, where the input signal is convolved with a standard HRIR. For instance, a sound source fixed at 60° in the median plane can be perceived at +45° to -30° degrees if the centre frequency of a notch filter is reduced from around 10 to around 6.3kHz[11].

However, each individual has a unique size and shape of different anatomical parts like the pinna, head and torso, and as a consequence each individual will also have a unique set of HRTFs[35]. Moreover, the depth of a notch is different from subject to subject[43]. In applications of binaural technology where non-individual HRTFs are used performance is degraded if the spectral characteristics of the filters do not match the listener's HRTFs. The result is front-back and up-down confusion.

There is considerable psychophysical and physiological evidence that the auditory system acts to smooth out variations in the spectral profile of a complex sound[14]. Filtering the HRTF to account for the variation in auditory filter bandwidth can result in the loss of a considerable amount of the fine structure in the HRTF. This suggests that, in so far as the very sharp peaks and notches in the HRTF occur within the bandwidth of single auditory filter, they are unlikely to be represented within the auditory nervous system. These features are called the perceptually salient features of the HRTF[14]. For this reason it is appropriate to further investigate the HRTFs after their smoothing at the final stage of the peripheral processor. Moreover from Figure 6 it appears that the current model will lose a lot of information above around 12kHz where some elevation cues are hidden.

5 CONCLUSIONS

This. paper has shown a number of ways in which the existing biologically inspired model may be extended to better match the auditory system, for instance by extending the range of frequencies over which it works and by incorporating HRTF spectral cues.

All the above-cited investigations were undertaken by visual observation of different HRTF databases [1, 32, 40] at different azimuth angles (θ') and by comparing the results with the literature. All HRTF databases reveal similar results regardless of the spherical coordinate system in which they have been measured and by adopting as a reference coordinate system the interaual-polar coordinate system, it has been shown that different azimuth angles (θ') form different spectral patterns in both ears. The possibility that the spectral patterns can be smoothed out by the auditory system makes necessary the further investigation of these in the context of the current auditory model.

6 FIGURES

a)

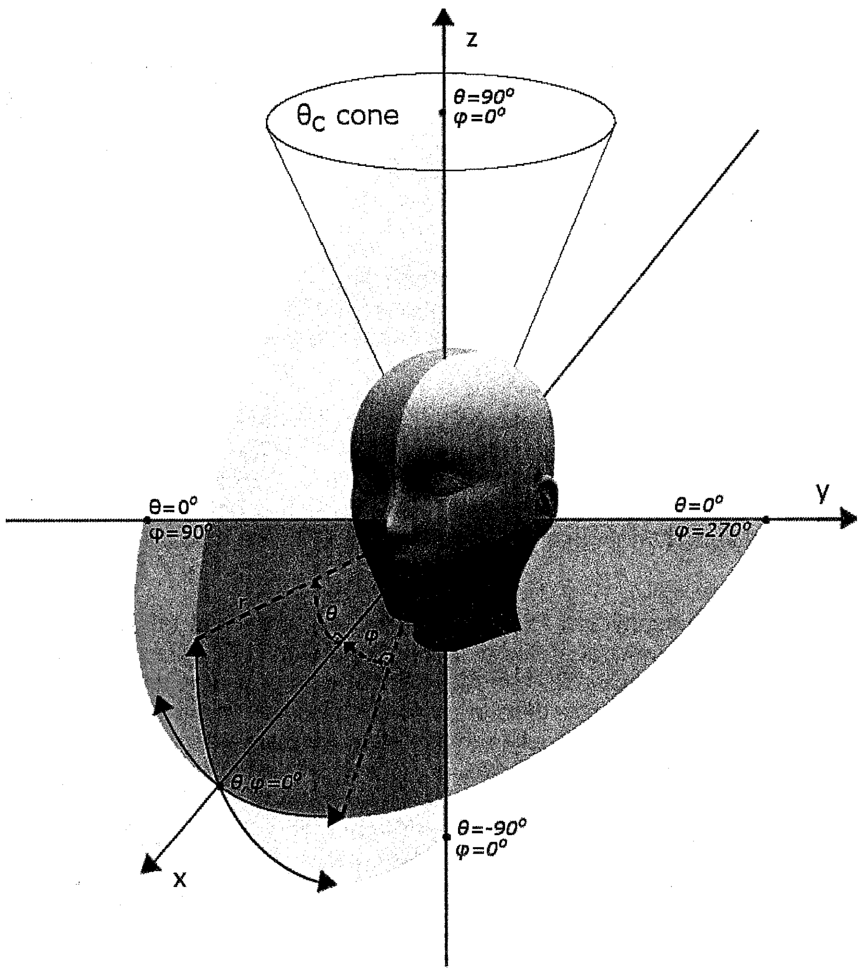

Figure 1 The conventional vertical-polar coordinate system is one of the most common head-related spherical coordinate systems defining the horizontal (xy) plane, the vertical (xz) (or median) plane, and distance, r.

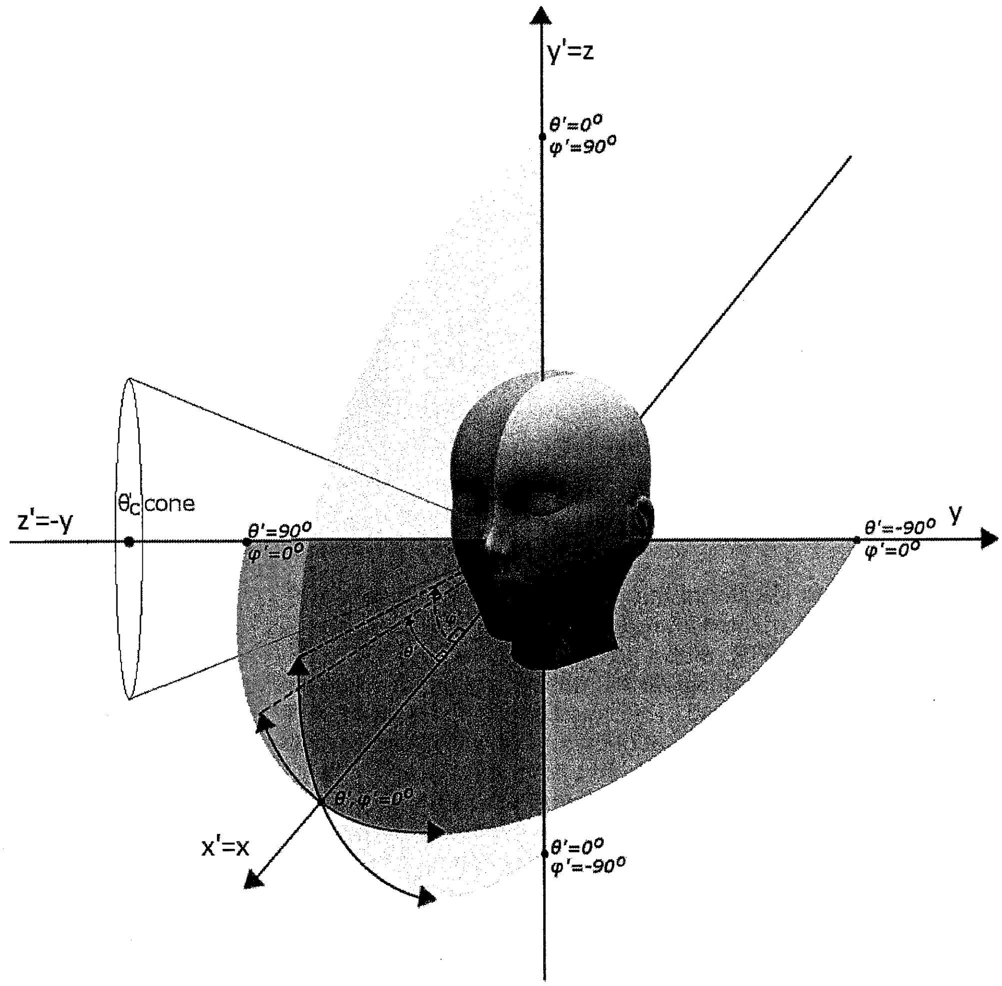

Figure 2 The interaural-polar coordinate system is an alternative head-related spherical coordinate system but instead of defining cones on the z axis (Figure 1), it defines them on the y axis. As a consequence for different azimuth angles (θ') different cones of confusions can be defined. Both systems are equivalent since the vertical-polar coordinate system can be transformed to the interaural-polar coordinate system with a 90° rotation of the axis about the x axis to the left. The interaural-coordinate system is the system that has been used in this paper.

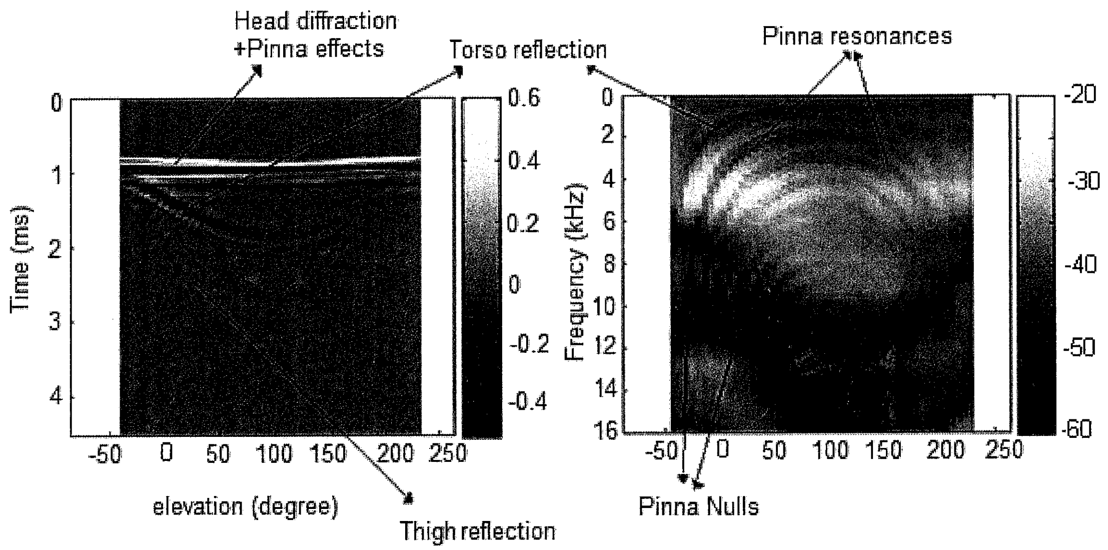

Figure 3 a) HRIRs, b) HRTFs of the right ear for azimuth angle (θ') 0° and elevation angles (φ') from -45° to +230.625°, of the 10th subject of the CIPIC database. The grey scale value represents in (a) the amplitude of HRIR, and in (b) the magnitude of HRTF in dB. In (a), the thigh reflection which appears in case of seated listeners doesn't appear in cases of KEMAR measurements where only the head and torso is used. This has been noticed in all HRTF measurements used in this paper. The torso reflection can be attributed mainly to shoulder reflections. For instance, at φ'=90° the torso reflection is delayed roughly 1ms from the onset of the HRIRs. Taking into account that the speed of sound in dry air at 20 °C is around 344 m/s, the distance that corresponds to that delay is 34cm. The half of this (17cm) is equal to the distance between the pinna and the shoulder of the current subject. Similar results can also be found in references [35, 39]

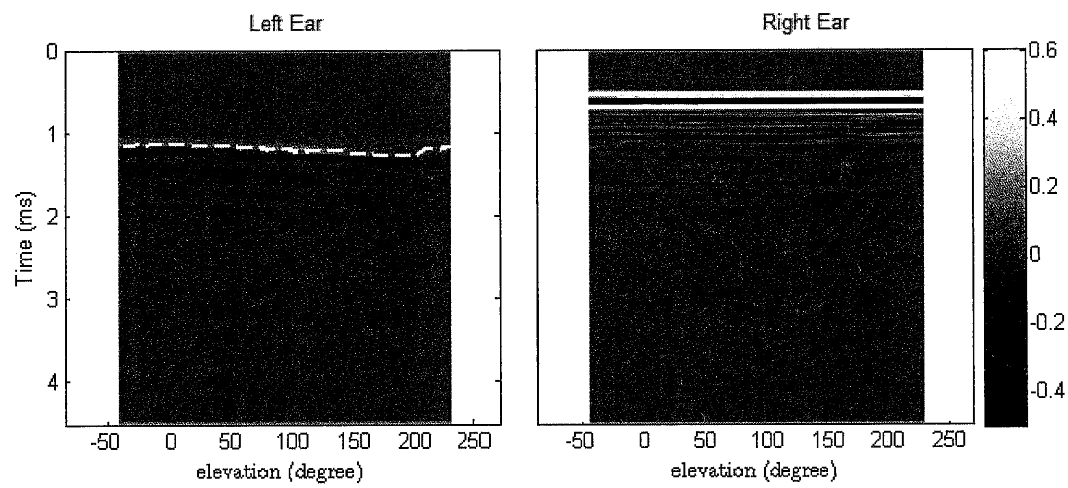

Figure 4 HRIRs of the left and right ear of the CIPIC database[1] for the 3rd subject for 80° azimuth angle (θ') on the subject's right side in the interaural-polar coordinate system. The horizontal line indicates the maximum value at each elevation angle (φ'). The grey scale value represents the amplitude of HRIRs.

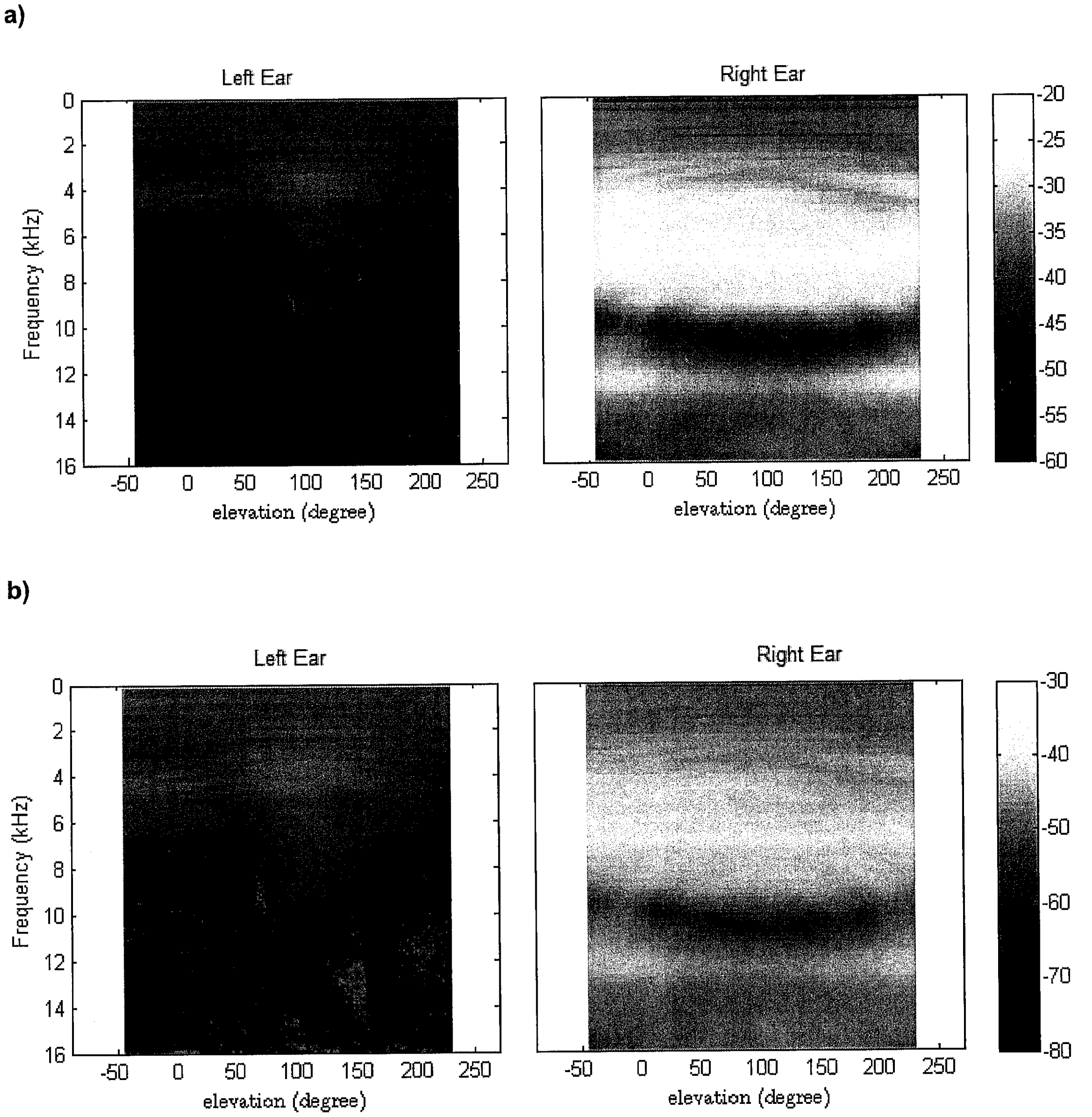

Figure 5 a) HRTFs of the CIPIC database[1] for the 3rd subject for 80° azimuth angle (θ') on the subject's right side in the interaural-polar coordinate system. b) The same as (a) with different range which was applied to better visualize the notches on the contralateral ear (left ear). The grey scale value represents the magnitude of the HRTF in dB.

a)

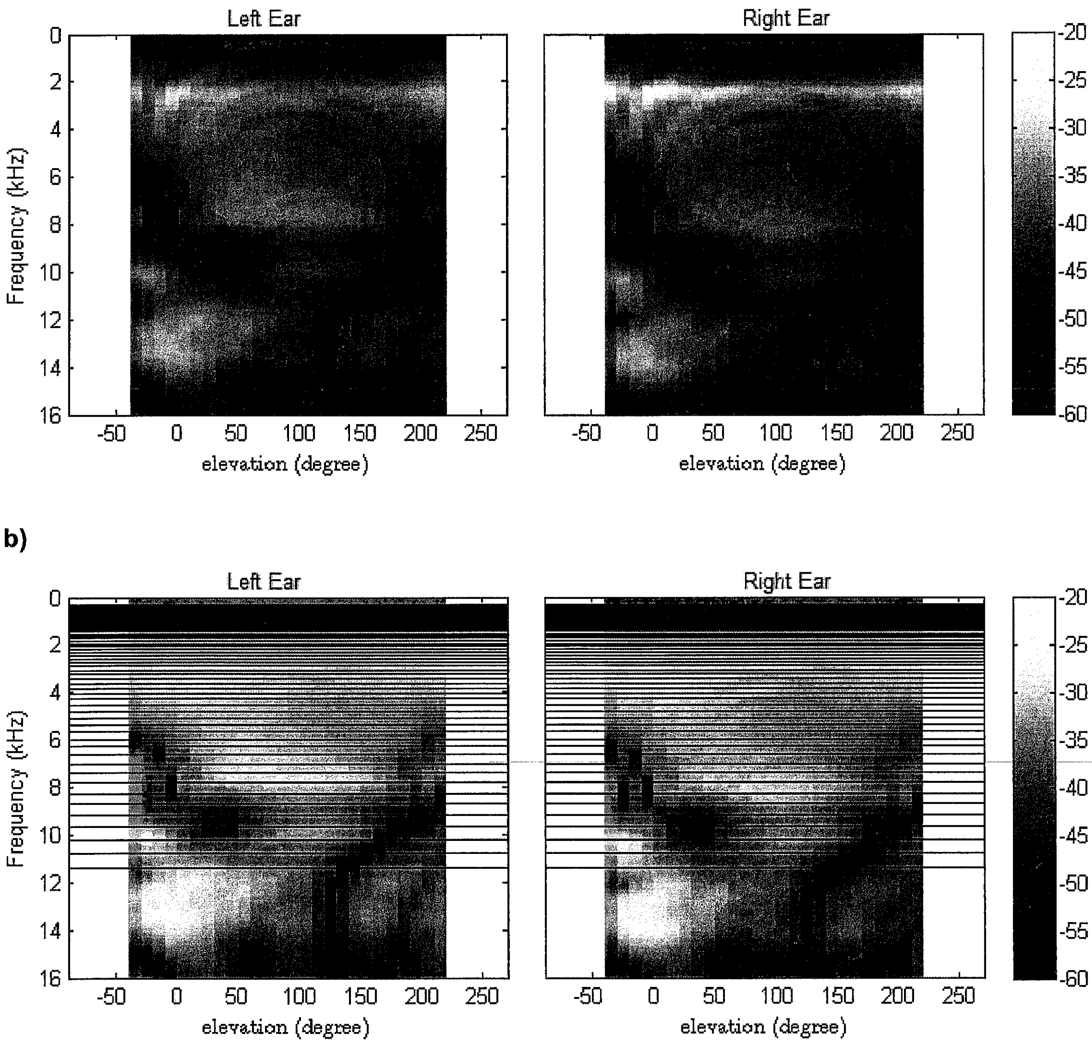

Figure 6 a) HRTFs of the ISVR[32] database for a KEMAR mannequin with large size pinna and open ear canal, in the median plane (θ' = 0°). (b) The same as a) but with the centre frequencies of the gammatone filters in the peripheral processor of the current model displayed as horizontal lines from 300Hz to 12kHz. The grey scale value represents the magnitude of the HRTF in dB.

7 REFERENCES

1 V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, 'The Cipic Hrtf Database', in *Applications of Signal Processing to Audio and Acoustics, 2001 IEEE Workshop on the*, 2001), pp. 99-102.

2 V. Ralph Algazi, Carlos Avendano, and Richard O. Duda, 'Elevation Localization and Head-Related Transfer Function Analysis at Low Frequencies', *The Journal of the Acoustical Society of America,* 109 (2001), 1110-22.

3 V. Ralph Algazi, Richard O. Duda, Ramani Duraiswami, Nail A. Gumerov, and Zhihui Tang, 'Approximating the Head-Related Transfer Function Using Simple Geometric Models of the Head and Torso', *The Journal of the Acoustical Society of America,* 112 (2002), 2053-64.

4 Jan W.H. Schnupp and Timothy P. Doubell Andrew J. King, 'The Shape of Ears to Come: Dynamic Coding of Auditory Space', *Trends in Cognitive Sciences,* 5 (2001), 261-2170.

5 Christian Feldbauer Anna Katharina Fuchs, Michael Stark, 'Monaural Sound Localization', *Signal Processing and Speech Communication Laboratory, Graz University of Technology, Austria* (2011).

6 Futoshi Asano, Yoiti Suzuki, and Toshio Sone, 'Role of Spectral Cues in Median Plane Localization', *The Journal of the Acoustical Society of America,* 88 (1990), 159-68.

7 C. Avendano, V. R. Algazi, and R. O. Duda, 'A Head-and-Torso Model for Low-Frequency Binaural Elevation Effects', in *Applications of Signal Processing to Audio and Acoustics, 1999 IEEE Workshop on*, 1999), pp. 179-82.

8 R. Batra, Kuwada, S., and Fitzpatrick, D. C, 'Sensitivity to Interaural Temporal Disparities of Low- and High-Frequency Neurons in the Superior Olivary Complex. Ii. Coincidence Detection', *Journal of Neurophysiology,* 78 (1997b), 1237-47.

9 R. Batra, Kuwada, S., and Fitzpatrick, D. C., 'Sensitivity to Interaural Temporal Disparities of Low- and High-Frequency Neurons in the Superior Olivary Complex. I. Heterogeneity of Responses', *Journal of Neurophysiology,* 78 (1997a), 1222–36.

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

11 P. Jeffrey Bloom, 'Creating Source Elevation Illusions by Spectral Manipulation', *Journal of the Audio Engineering Society,* 25 (1977), 560-65.

12 James C. Boudreau, and Chiyeko Tsuchitani, 'Binaural Interaction in the Cat Superior Olive S Segment', *The Journal of the Acoustical Society of America,* 42 (1967), 1207-07.

13 Jeroen; van de Par Breebaart, Steven; Kohlrausch, Armin, 'Binaural Processing Model Based on Contralateral Inhibition. I. Model Structure', *The Journal of the Acoustical Society of America,* 110 (2001), 1074-88.

14 S. Carlille, and D. Pralong, 'The Location-Dependent Nature of Perceptually Salient Features of the Human Head-Related Transfer Functions', *The Journal of the Acoustical Society of America,* 95 (1994), 3445-59.

15 Dan Jiang David McAlpine, Trevor M. Shackleton, and Alan R. Palmer, 'Convergent Input from Brainstem Coincidence Detectors onto Delay-Sensitive Neurons in the Inferior Colliculus', *The Journal of Neuroscience,* 18 (1998), 6026-39.

16 N. I. Durlach, 'Equalization and Cancellation Theory of Binaural Masking-Level Differences', *The Journal of the Acoustical Society of America,* 35 (1963), 1206-18.

17 H. Geoffrey Fisher, and Sanford J. Freedman, 'Localization of Sound During Simulated Unilateral Conductive Hearing Loss', *Acta oto-laryngologica,* 66 (1968), 213-20.

18 H. Geoffrey;Freedman Fisher, Sanford J., 'The Role of the Pinna in Auditory Localization', *Journal of Auditory Research,* 8 (1968), 15-26.

19 Mark B. Gardner, and Robert S. Gardner, 'Problem of Localization in the Median Plane: Effect of Pinnae Cavity Occlusion', *The Journal of the Acoustical Society of America,* 53 (1973), 400-08.

20 Lloyd A. Jeffress, 'A Place Theory of Sound Localization', *Journal of Comparative and Physiological Psychology,* 41 (1948), 35-39.

21 David M. Green John C. Middlebrooks, 'Sound Localization by Human Listeners', *Annual Review of Psychology,* 42 (1991), 24.

22 P. X. Joris, 'Envelope Coding in the Lateral Superior Olive. Ii. Characteristic Delays and Comparison with Responses in the Medial Superior Olive', *Journal of Neurophysiology,* 76 (1996), 2137–56.

23 P. X. Joris, and Yin, T. C. T., 'Envelope Coding in the Lateral Superior Olive. I. Sensitivity to Interaural Time Differences', *Journal of Neurophysiology,* 73 (1995), 1043-62.

24 Erno H. A. Langendijk, and Adelbert W. Bronkhorst, 'Contribution of Spectral Cues to Human Sound Localization', *The Journal of the Acoustical Society of America,* 112 (2002), 1583-96.

25 Enrique A.; Meddis Lopez-Poveda, Ray 'A Physical Model of Sound Diffraction and Reflections in the Human Concha', *Acoustical Society of America,* 100 (1996).

26 J. C. Middlebrooks, & Green, D. M., 'Sound Localization by Human Listeners', *Annual Review of Psychology,* 42 (1991), 135–59.

27 A. W. Mills, *Auditory Localization*. Vol. 2, *Foundations of Modern Auditory Theor*Academic, New York, 1972).

28 Henrik; Toledo Møller, Daniela, 'The Role of Spectral Features in Sound Localization', *Audio Engineering Society Convention 124/7450* (2008).

29 Brian C. J. Moore, Simon R. Oldfield, and Gary J. Dooley, 'Detection and Discrimination of Spectral Peaks and Notches at 1 and 8 Khz', *The Journal of the Acoustical Society of America,* 85 (1989), 820-36.

30 Brian C.J. Moore, *An Introduction to the Psychology of Hearing*. 5th ednAcademic Press, 2003).

31 Park Munhum, 'Models of Binaural Hearing for Sound Lateralisation and Localisation', University of Southampton, 2007).

32 M. Park P. Mannerheim, T. Papadopoulos and P.A. Nelson, 'The Measurement of a Database of Head Related Transfer Functions', *ISVR Contract Report , E. J. Richards Library, University of Southampton, Institute of Sound and Vibration Research, Fluid Dynamics and Acoustics Group,* No. 04-07 (2004).

33 Munhum; Nelson Park, Philip A.; Kang, Kyeongok, 'A Model of Sound Localisation Applied to the Evaluation of Systems for Stereophony', *Acta Acustic United with Acustica,* 94 (2008), 825-39.

34 Ville Pulkki, and Toni Hirvonen, 'Functional Count-Comparison Model for Binaural Decoding', *Acta Acustica united with Acustica,* 95 (2009), 883-900.

35 Vikas C. Raykar, Ramani Duraiswami, and B. Yegnanarayana, 'Extracting the Frequencies of the Pinna Spectral Notches in Measured Head Related Impulse Responses', *The Journal of the Acoustical Society of America,* 118 (2005), 364-74.

36 T. C. Yin S. Kuwada, J. Syka, T. J. Buunen, and R. E. Wickesberg, 'Binaural Interaction in Low-Frequency Neurons in Inferior Colliculus of the Cat. Iv. Comparison of Monaural and Binaural Response Properties', *Journal of Neurophysiology,* 51 (1984), 1306-25.

37 C. L. Searle, L. D. Braida, M. F. Davis, and H. S. Colburn, 'Model for Auditory Localization', *The Journal of the Acoustical Society of America,* 60 (1976), 1164-75.

38 Edgar A. G. Shaw, 'Acoustical Characteristics of the Outer Ear', in *Encyclopedia of Acoustics*John Wiley & Sons, Inc., 2007), pp. 1325-35.

39 Ramani Duraiswami Vikas C. Raykar, Larry Davis, B. Yegnanarayana, 'Extracting Significant Features from the Hrtf', *Proceedings of the 2003 International Conference on Auditory Display, Boston, MA, USA, July 6-9, 2003* (2003).

40 O Warusfel, 'Ircam/Akg Listen Hrtf Database'2003) <http://recherche.ircam.fr/equipes/salles/listen/index.html>2012].

41 Frederic L. Wightman, and Doris J. Kistler, 'Monaural Sound Localization Revisited', *The Journal of the Acoustical Society of America,* 101 (1997), 1050-63.

42 V. and Eggert Willert, J. and Adamy, J. and Stahl, R. and Korner, E., 'A Probabilistic Model for Binaural Sound Localization', *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on,* 36 (2006), 982 -94.

43 BoSun Xie, XiaoLi Zhong, Dan Rao, and ZhiQiang Liang, 'Head-Related Transfer Function Database and Its Analyses', *Science in China Series G: Physics Mechanics and Astronomy,* 50 (2007), 267-80.

44 T. C. T. Yin, and J. C. K. Chan, 'Interaural Time Sensitivity in Medial Superior Olive of Cat', *Journal of Neurophysiology,* 64 (1990), 465–88.