Further Developing 3D Panning for Headphones

Sally-anne Kellaway S430029204
Digital Audio Systems 2013 DESC9115
Submission for Written Review 2

Product proposition

Ongoing challenges greet both academic and private researchers interested in developing sound specialization techniques that are applicable to a wide range of listeners. Raykar, Duraiswami and Yegnanarayana (2005) raise several prevalent methods of achieving spatialzation for digital audio, inclusive of HRTF measurement, database matching, numerical modelling and frequency scaling to attempt to match non-individualised HRTFs to an individual. Spatialization of audio developed for headphones is applicable to many purposes – the growing penetration of virtual reality technology, video game environments and application to physical artistic installations, as well as teleconferences demonstrate the variety of environments that this technology could benefit. This project aims to simplify the spatial cues driven by the structure of the human pinna for application to teleconferencing.

Product Goal

This product aims to provide the listener the ability to detect elevation difference between two or more sources. Previous projects by the author have discussed in deep detail the development and execution of the processing to execute spatialization effects on the azimuth plane. As demonstrated in Figure 1, the role of previous developments falls within the following overall system diagram.

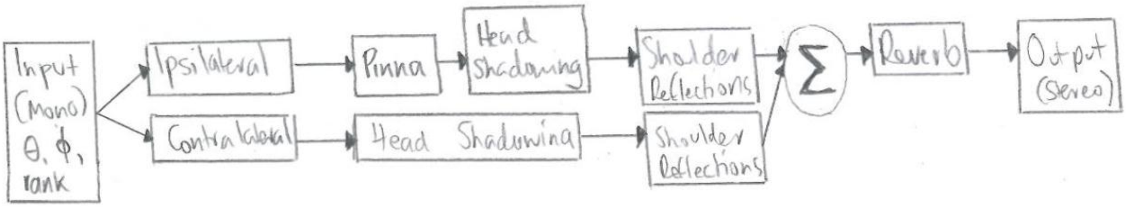

Figure 1 – System diagram over the overall Spatialization for Headphones product.

As raised in the author's previous works, and supported by Raykar, Duraiswami and Yegnanarayana, Interaural Time and Level differences are primary cues for localisation on the azimuth plane (2005). Models for these systems leave the vertical plane without discussion or solution. This project aims to present a solution, which provides elevation cues in a way which can be assimilated with the existing system without introducing detrimental amounts of extra processing.

Execution

Background

The system, which this module is to fit with, operates on the inputted audio within the time domain. For the more simplistic systems of Head Shadowing/ITD (HSH) and Shoulder Reflections (SHR) existing models are well established (Algazi et al, 2002; Duda and Martens, 1998) – more specifically as single delays in the time domain (DAFX) Inter-aural Time Differences (ITD), Shoulder Reflection (SHR) have existing models. Approximating the effect of the pinna module to *n* number of delays in the time domain is thought to discard significant spectral information, as discussed by Han in great length. “No part of the spatial ranges covered by the measurements is devoid of meaningful spectral features.” (p 25 1994). Han then concludes that the influence of the pinna operates in the spectral domain as opposed to the time domain. Applying this approach to this product, we seek to summarise the major spectral cues related to elevation to a simplified pole-zero system.

Initial survey of related literature uncovers a range of disagreement over the number of reflections that are required to achieve adequate replication of the effect of the pinna, and whether resonances are required. In Barreto and Gupta (2003), there is discussion of Batteau’s three-path acoustic coupler, where two delays are calculated with the following;

$$\tau_1 = (\frac{S_1}{c})(1 + cos\theta)$$
$$\tau_2 = \left(2\frac{S_1}{c}\right) + (\frac{S_2}{c})(1 + sin\phi)$$

Where S_1 and S_2 are the separation between the two holes in the coupler, and a third reference hole, c is the speed of sound, θ is the azimuth angle and ϕ the elevation angle. This model only presents one delay, which is reactive to elevation, when simple inspection of a HRTF function (for example, Han’s Pinna-only HRTFs, 1994) demonstrate that there are much more complicated spectral effects taking place.

Barreto and Gupta then compare this approach with a five-delay model from Brown and Duda.

$$\tau_k(\theta,\phi) = A_k \cos\left(\frac{\theta}{2}\right)\sin\left(D_k(90° - \phi)\right) + B_k,$$
$$\text{for } k = 1 \ldots 5$$

Although there are 5 iteration-based delays within this model, the constants A_k, B_k and D_k are defined and adjusted in an *ad hoc* manner and hence highlights the need to establish a fully systematic approach to modeling the effect of the pinna. While having variables that assist in customizing the result to the listener’s perception of elevation is valuable in environments where either the appropriate interface, or an “editor” with the required knowledge is present, any of the potential uses listed previously for this product do not allow such luxuries. A deeper understanding of how the features of the result relate to the anthropological measurement of the listener's pinna is critical for this product.

Han's approach utilizes the analysis of HRTFs recorded from a dummy head to isolate the impact of not only the pinna, but two major areas of the pinna – the concha and the helix to concha area.

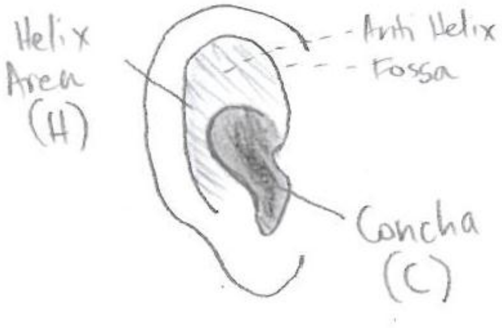

Figure 2 – A descriptive diagram of the pinna, featuring the concha (C) and the helix to concha area (H).

Anthropological Description of Pinna Effect

Han uses a range of measured HRTFs from the KEMAR dummy to isolate the functional purposes of these two areas – C is more active when the source is above the listener, and H when the sound from below. Also stated are simple equations for frequency-domain calculation of the spectral notches – immediately annotated with claim that measured results varied from calculated results as they always will due to the unique nature of the pinna on a person-to-person basis.

$$notch\ f = \frac{3T}{2}$$

where T=1/f

However, Han demonstrates with far reaching clarity that not only is the concha solely responsible for splitting the first arriving sound to denote elevation (a delay of 46.87 µs to 62.5 µs was measured for KEMAR between 0 and 60° elevation), but it also "appears to play the major role in horizontal plan localization" (p 27).

The strong reflection provided by the concha creates destructive effects, which, largely conform to the mostly linear movement of a singular notch which sits at approximately 9.5 kHz and deepens from appx -2 dB at 60° elevation to -10dB at 0°. From here, sources below will cause a double notch pattern between -10° and -30°, moving closer to 7kHz, and then wide double-notches between -40° and -60° elevation. A secondary notch moves from 15kHz to 12 kHz within the same elevation range. This is also accompanied by peak that sits between these notches, as well as additional details, which would invariably differ depending on the subject measured.

Hebrank and Wright's subjectively derived model for discerning elevation and front-back cues on the median plane are compared with Han's model. The bands defined by Hebrank and Wright that the subject group nominated as giving the effect of elevation, overhead and rear placement largely support Han's results, and further support the summary that will be implemented for this project. Particular interest lies within the specification of a band which has marked effect on front-back shadowing as it is not specifically discussed within the body of Han's research, or the collective literature pool used in the development of this product. A notch develops as sources shift between 90° and 180° azimuth (p 33), which appears to have a more dramatic effect if the source is at a negative elevation. Information presented in the sources do not disclose a conclusive enough set of HRTFs to determine the exact movements of this notch, however, this set of information is adequate to design the filter for testing and manipulation.

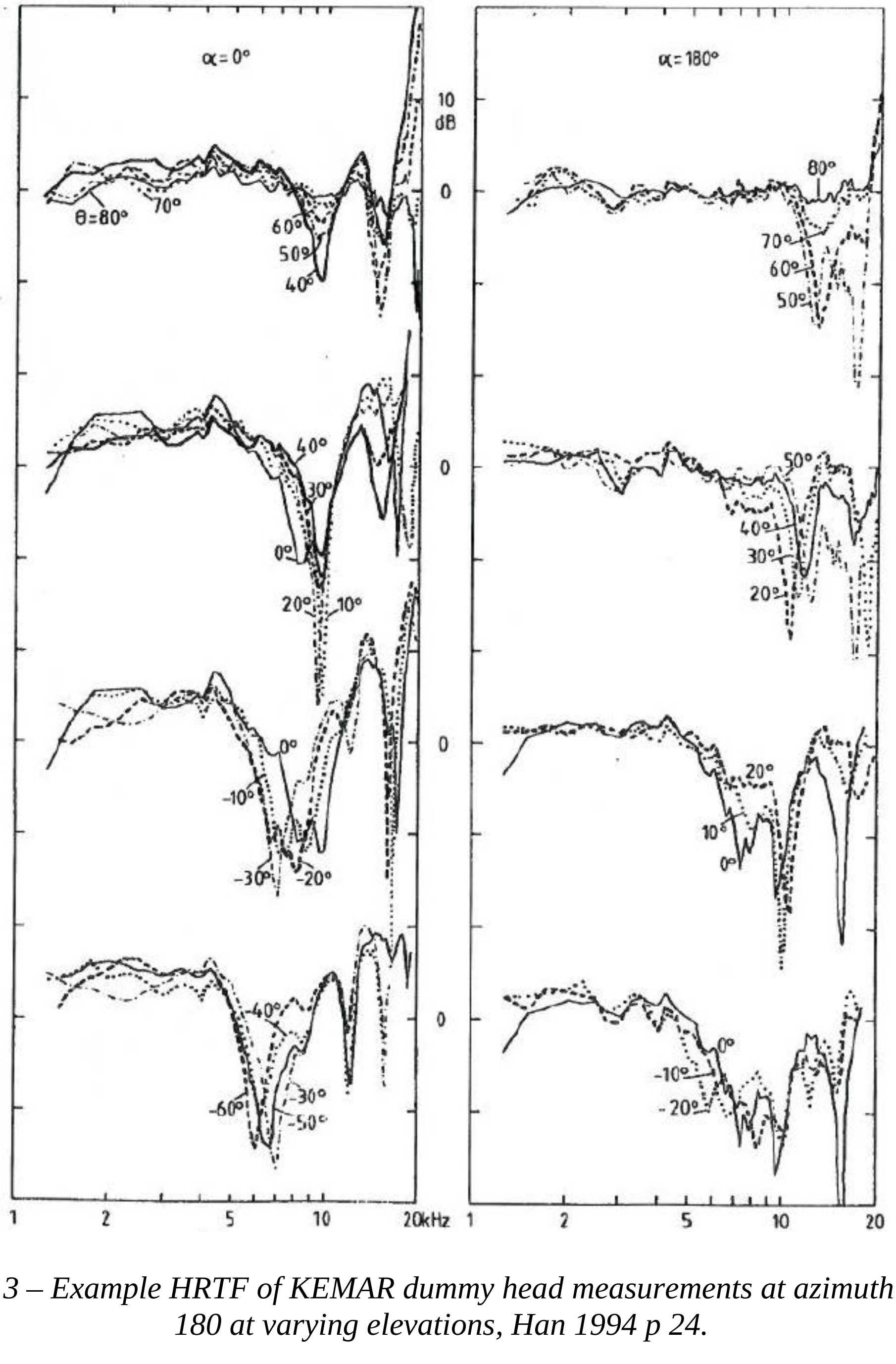

Figure 3 – Example HRTF of KEMAR dummy head measurements at azimuth 0 and 180 at varying elevations, Han 1994 p 24.

These descriptions are therefore summarised into the following table, which describes the behaviour that the pole-zero system should have moving through elevation and azimuth.

ϕ	θ = 0°			θ = 45°		θ = 90°		
	L	V	←W	L	V	L	V	P
90°	🟥	9.5		🟥		🟥	12.75	
60°		↓		🟥	9.5	🟥	↓	
30°		↓		🟥	↓		12.5	9-8
0°		→	8-7	🟥	10.5		←	
-30°		9.75			←		←	
-60°	7-6	←		🟥	6.5		←	6

ϕ	θ = 135°				θ = 180°		
	L	←V	V2	V3	V	←W	V
90°	🟥				↓		
60°		12.75	6		↓		
30°		11	🟪	16	←		16
0°		🟪	🟪	↓	←	7.25	↓
-30°		8.5	4.25	↓		🟦	↓
-60°						5.75	

Tables 1 and 2 – Movement of peak and notch filters by azimuth and elevation. Adpated from Han 1994, p 26)

Legend – Tables 1 and 2	
L	Level increase
V	Valley
W	Double Valley
P	Peak
θ	Theta, Azimuth
ϕ	Phi, Elevation
□	Broadening of valley
□	Narrowing of valley
□	Deepening of Valley
9-8	9kHz – 8kHz frequency values

Application to software such as MATLAB is anticipated for this project. It has been previously nominated that a pole-zero filter will suit the manipulations required by this system. Tables 1 and 2 will assist in creating parameters to begin testing the pinna model, with a few simplifications, which can be made at this stage as per the following;

- Level differences can be omitted as this effect is implemented by an ITD/Head Shadowing function
- Implementing the tabled variables to an array or lookup table
- Implementing a conditional variable test for the azimuth and elevation declared by the user to discover which azimuth and elevation manipulations need to be implemented by the pole-zero filters
- Applying the pole-zero filters to the input of the system and outputting one (ipsilateral) file to feed in to the rest of the system. The contralateral ear is not tested and due to the lack of direct sound, is unlikely to experience similar depth of pinna effects

The pinna model system will hence apply the following;

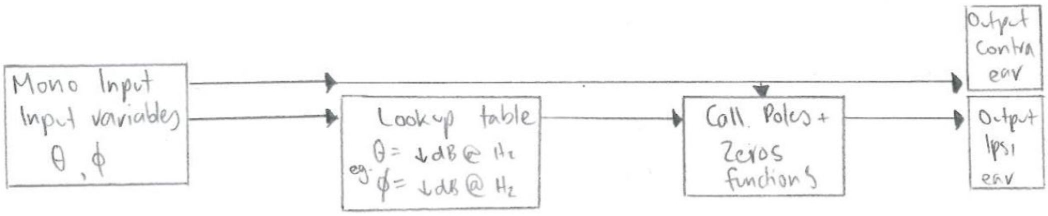

Figure 4 – System diagram of the Pinna Modelling element.

Conclusion - Expected Results and Future Development

This project has proposed the development of a product for the purpose of application to teleconferencing. With the background provided, a suitable level of data has been discussed to support the proposed approach. After the proposed system has been constructed, it is anticipated that there should be testing with multiple subjects in order to test the suitability of the derived spectral filters to a general listening audience. It is further anticipated that there may need to be allowances made to develop additional curves for different ear shapes – as in Raykar, Duraiswami and Yegnanarayana (2005), this can be made possible by analysing additional HRTFs from the CIPIC database. One example of a major difference that might need to be catered for is the difference between notch patterns for individuals for whom the concha is the primary source of reflection and resonance (such as the implementation discussed through the majority of this proposal) and those for whom the helix to concha area is the primary source of reflection and resonance (as raised in Algazi, Duda and Satarzadeh, 2007).

It is anticipated that strong elements of this implementation will be the manner by which the reflections provided will intrinsically relate to the elevation cues (in comparison to approaches such at the Batteau), and additionally help support the azimuth filtering effects in the Cone of Confusion (which is a limitation of ITD and ILD-only systems). This approach is also anticipated to be more accurate than the time-domain approach raised in the DAFX text from Rochesso (2002) as time-domain solutions are like to oversimplify the complex spectral information that the pinna provides (for example, the combination of deep valley and double valley at 0° azimuth).

References and Bibliography

Algazi VR, Duda RO, Duraiswami R, Gumerov NA, Tang Z, 2002, ‘Approximating the Head-Related Transfer Function Using Simple Geometric Models of the Head and Torso’, *Journal of the Acoustical Society of America, 112* (5) 2053-2064.

Algazi VR, Duda RO and Satarzadeh P, 2007, ‘Physical and Filter Pinna Models Based on Anthropometry’ Proc. 2007 AES Convention, Vienna, Austria, May 2007.

Barreto A and Gupta N 2003, ‘Dynamic Modeling of the Pinna for Audio Spatialization’ Proc. 2003 WSEAS, Puerto De La Cruz, Tenerife, Canary Islands, Spain., December 2003.

Brown, C, 1996, ‘Modeling the Elevation Characteristics of the Head-Related Impulse response’, M. thesis, San Jose State University, San Jose.

Duda RO, and Martens WL, 1998, ‘Range Dependance of the Response of a Spherical Head Model’, *Journal of the Acoustical Society of America, 104* (5) 3048-3058).

Han, H, 1994, ‘Measuring a Dummy Head in Search of Pinna Cues’, *Journal of the Audio Engineering Society, 42*(1/2) 15-37).

Han, H, 1994, chart, in ‘Measuring a Dummy Head in Search of Pinna Cues’, *Journal of the Audio Engineering Society, 42*(1/2) 23.

Martens, W (2013) Class notes for *Week 10*. Retrieved June 3, from http://web.arch.usyd.edu.au/~wmar0109/DESC9115/slide_sets/W8_IIR_Z_Handout.pdf

Raykar VC, Duraiswami R and Yegnanarayana B, 2005, ‘Extracting the frequencies of the pinne spectral notches in measured head related impulse responses’ *Journal of the Acoustical Society of America, 118* (1), 364-374.

Rocchesso, D. (2002), 'Spatial Effects', in U Zolzer (ed.), *DAFX – Digital Audio Effects*, John Wiley & Sons Ltd, Stafford, QLD.