STRUCTURAL MODELLING OF THE HRTF: A BALANCED MODEL TRUNCATION OF THE INTERAURAL LEVEL DIFFERENCE

CHRISTOPHER BUCHANAN ACOUSTICS & AUDIO GROUP, UNIVERSITY OF EDINBURGH

OVERVIEW

- The HRTF (Head-Related Transfer Function) is key in the synthesis of 3D audio. It is a complex-valued function of 4 variables: azimuth **(**θ**)**, elevation, range **(**ρ**)** and frequency, and contains the information our brain requires for determining sound source location.
- Real-time rendering of 3D audio needs to include full consideration of the physics governing sound wave propagation in both the **near** (within $\approx$ 1m of the head) and **far** field.
- The aim of this project was to derive and implement a modern customizable structural model of the the Interaural Level Difference (ILD), a binaural cue essential in the **lateral** localization of sound.

THE SPHERICAL TRANSFER FUNCTION (STF)

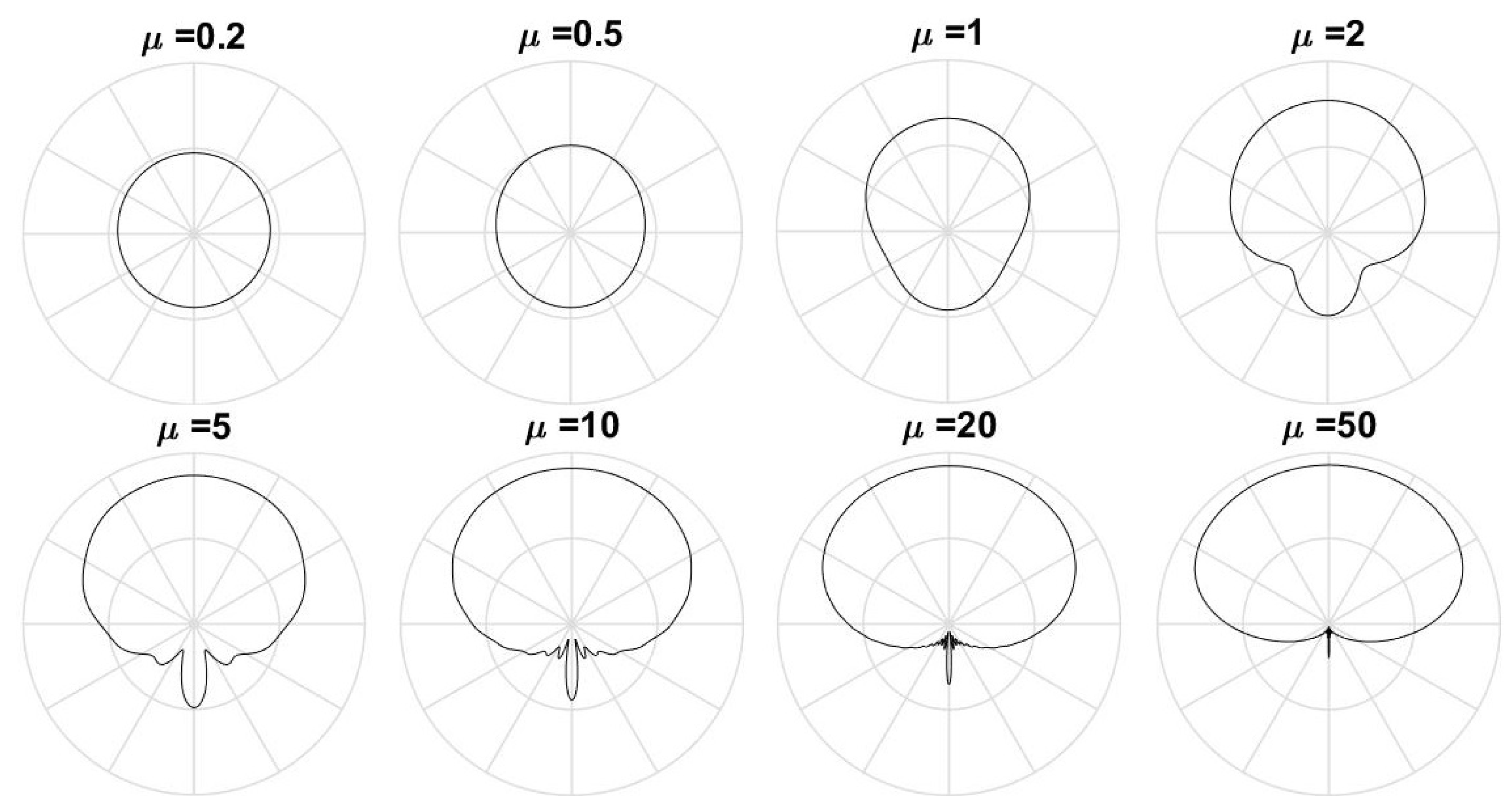
In most real-time simulations involving synthesis of HRTFs, the replacement of the human head by a rigid sphere of equal average radius is a well-motivated one. Whilst in general the HRTF of any given individual cannot be analytically obtained, the STF can be expressed as an infinite series:

$$H(\rho,\mu,\theta) = -\frac{\rho}{\mu}e^{-i\mu\rho}\Psi \quad (1)$$

where

$$\Psi(\rho,\mu,\theta) = \sum_{m=0}^{\infty}(2m+1)P_m(\cos\theta)\frac{h_m(\mu\rho)}{h'_m(\mu)}$$

with h_m being the m^{th} order spherical Hankel function, and h'_m its derivative. P_m represents a Legendre polynomial of order m, and normalized variables $\rho = r/a$ and $\mu = ka$ represent range and frequency respectively (a = sphere radius, r = source distance).



Polar plots of the absolute magnitude of the STF for $\rho = 32$.

A closed-form analytical solution does not exist for the STF, but recursive numerical methods such as Duda & Martens [4] perform well as an appropriate representation.

REFERENCES

1. Simone Spagnol, Michele Geronazzo, and Federico Avanzini. Hearing distance: A low-cost model for near-field binaural effects. 20th European Signal Processing Conference (EUSIPCO 2012), pages 2030-2034, 2012.
2. S.Spagnol and F.Avanzini. Real-time binaural audio rendering in the near field. Proc. 6th Int. Conf. Sound and Music Computing (SMC09), pages 201-206, 2009.
3. Bartlomiej Beliczynski, Izzet Kale, and Gerald D. Cain. Approximation of fir by iir digital filters: An algorithm based on balanced model reduction. IEEE Transactions on Signal Processing, 40(3):158-173, 1992.
4. Richard O. Duda and William L. Martens. Range dependence of the response of a spherical head model. J. Acoust. Soc. Am., 104(5):3048-3058, 1998.

BALANCED MODEL TRUNCATION

To obtain an M^{th} order FIR filter replicating the STF H at a given θ and ρ, we first sample it at M suitably distributed frequencies within the Nyquist range before applying an inverse Discrete Fourier Transform with Hanning windowing. The choice of M and the distribution of sample frequencies can be parameterized to suit computational cost and quality requirements. Each of these FIRs represents a point within a (θ,ρ) grid, and in general each transfer function $c_0 + c_1z^{-1} + \cdots + c_Mz^{-M}$ can also be written as a set of state-space difference equations:

$$x(k+1) = Ax(k) + Bu(k)$$
$$y(k) = Cx(k) + Du(k)$$

where

$$A = \begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \\ & & \cdots & & \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad C = \begin{pmatrix} c_1 & c_2 & \cdots & c_M \end{pmatrix}, \quad D = c_0 \quad (2)$$

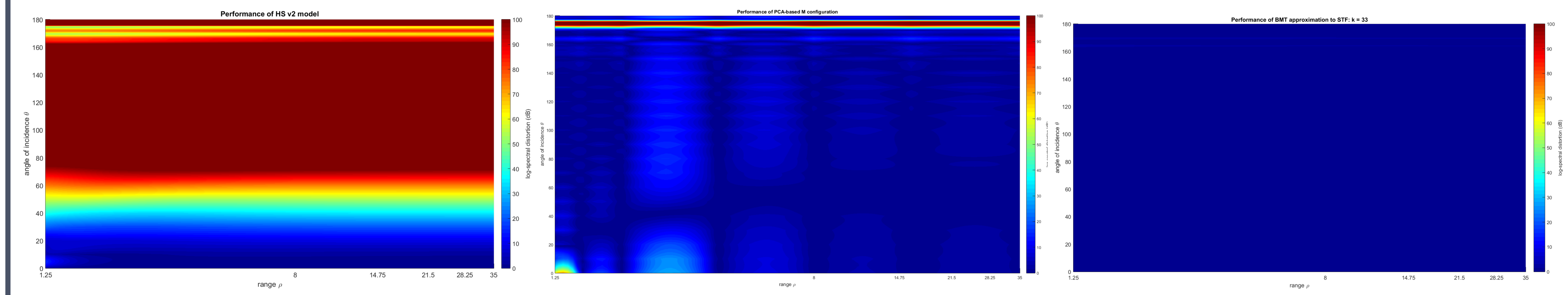
A useful advantage of considering an FIR in state-space system form lies in its associated Hankel matrix, which comprises only of the coefficients $c_0, c_1, \cdots, c_M$. A convenient measure of the impact of model truncation is encoded within the eigenvalues of this matrix. By analysing these values against their index of decreasing magnitude, one can determine a sensible point at which to **reduce** the system, obtaining a value for k in doing so for application of the following theorem [3]:

Theorem 1. *If the Hankel matrix of an M^{th} order FIR system (A, B, C) is diagonalizable, then a k^{th} order reduced balanced system is input/out equivalent to the system (A_k, B_k, C_k) where*

$$A_k = V(2:M,1:k)'V(1:M-1,1:k)$$
$$B_k = V(1,1:k)' \quad (3)$$
$$C_k = CV(1:n,1:k)$$

where $V(i:j,k:m)$ denotes an extraction of rows i to j and columns k to m of V (a matrix whose columns are eigenvectors of the Hankel matrix sorted 'largest' to 'smallest').

EVALUATION



Log-spectral distortion (LSD) plots for the Near-Field shelving filter model (left) [1], Principal Components Analysis (middle) [2] and BMT (right), as a function of θ and ρ.

The accuracy of approximation can be attested through LSD, with the numerically obtained STF as our 'benchmark'.

1. For the NFTF-shelving-filter approach [1], the cascading of simple filter topologies ensures a cheap and versatile system, but accuracy beyond $\theta = 60°$ suffers broadly
2. Whilst various components of the PCA method [2] can be generated offline and is of purely statistical derivation, the reliance on Yule-Walker algorithm heavily restricts flexibility. The dreaded near-contralateral region ($\theta \approx 180°$) also causes issues
3. The BMT-based approach exhibits noticeable accuracy improvements across all angles and ranges

The BMT methodology described here is of stable, analytical derivation and allows wide scope for parameterisation at both the FIR and IIR modelling stages. Through careful handling of its resultant IIR filterbank, it could benefit any structural model of the HRTF.