Audio Engineering Society

Convention Paper

Presented at the 125th Convention
2008 October 2–5 San Francisco, CA, USA

On the Minimum-Phase Nature of Head-Related Transfer Functions

Juhan Nam[1], Miriam A. Kolar[1], and Jonathan S. Abel[1]

[1] *Center for Computer Research in Music and Acoustics (CCRMA), Stanford Univ., Stanford, CA, 94305, USA*

Correspondence should be addressed to Juhan Nam (juhan@ccrma.stanford.edu)

ABSTRACT

For binaural synthesis, head-related transfer functions (HRTFs) are commonly implemented as pure delays followed by minimum-phase systems. Here, the minimum-phase nature of HRTFs is studied. The cross-coherence between minimum-phase and unprocessed measured HRTFs was seen to be greater than 0.9 for a vast majority of the HRTFs, and was rarely below 0.8. Non-minimum-phase filter components resulting in reduced cross-coherence appeared in frontal and ipsilateral directions. The excess group delay indicates that these non-minimum-phase components are associated with regions of moderate HRTF energy. Other regions of excess phase correspond to high-frequency spectral nulls, and have little effect on cross-coherence.

1. INTRODUCTION

A head-related transfer function (HRTF) is typically implemented as the cascade of a pure delay and a minimum-phase filter as shown in Fig.1 [1, 2]. Using this model reduces the cost of binaural synthesis by shortening the length of FIR filters [3] and reducing the number of components needed in linear decomposition methods [4]. In addition, smooth interpolation of head-related impulse responses (HRIR) at discrete directions, necessary for simulation of a moving sound source, highly benefits from the minimum-phase representation [5].

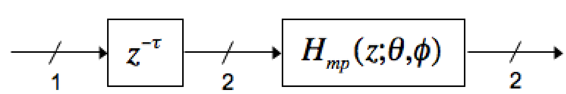


Fig. 1: The minimum-phase model of HRTF

Studies to understand HRTFs in terms of minimum-phase functions have been carried out by many authors. Mehrgardt and Mellet [6] found HRTFs to be nearly minimum-phase up to 10 kHz. Wightman and Kistler [1], in pursuing a principal component analysis, assumed that the measured HRTFs are minimum-phase functions with interaural phase differences approximated with a pure delay. Kulkarni et al. [7, 8] examined the validity of the the minimum-phase model by subjective testing with modified phase spectra including minimum-phase plus delay, linear-phase, and reversed-phase plus delay. They also analyzed the similarity between an HRTF and its minimum-phase component. Plogsties et al. [5] tested audibility of allpass HRTF components. They showed that the allpass components can be removed without audible consequences, except in certain cases which can be ignored if the allphase components are replaced with differences of group delay at DC.

Although these studies employed different methods to estimate interaural time difference (ITD), the common conclusion of subjective experiments is that the minimum-phase model is indistinguishable from the HRTFs on which it was based if the appropriate ITD is applied. This brings up the question of whether this conclusion results from human hearing insensitivity to the non-minimum-phase components, or if HRTFs are themselves essentially minimum-phase. In this paper we examine HRTF phase, and show that HRTFs are essentially minimum-phase.

As a means of evaluating the minimum-phase content of HRTFs, cross-coherences between HRTFs and their minimum phase versions are computed. The idea is that if the HRTF were a pure delay followed by a minimum-phase system, the cross-coherence would be 1.0, since the filters are time-shifted versions of each other. We refer to this cross-coherence between the original and minimum-phase HRTFs as the *retained coherence*; the majority of HRTFs have retained coherence above 0.9, indicating that the bulk of HRTF energy is minimum-phase. Kulkarni et al. [7, 8] computed the retained coherence for two subjects' HRTFs and noted that the vast majority of the HRTFs have retained coherence above 0.9. It will be further explored in this paper.

Another approach to evaluate the minimum-phase content of HRTFs is to examine the excess group delay. When the excess group delay is a constant function of frequency, the HRTF is a pure delay followed by a minimum-phase filter. In measured HRTFs, two distinct types of non-constant excess group delay were found—one associated with regions of moderate spectral energy and another associated with isolated spectral nulls, both typically above 10 kHz. In the following, we explain HRTF retained coherence and then the HRTF group delay.

2. HRTF COHERENCE

We begin by introducing some notation. The HRTF at frequency ω and azimuth and elevation $(\theta,\ \phi)$ is denoted by $H(\omega;\theta,\phi)$. Suppressing the arrival direction dependence, the HRTF may be written in magnitude-phase form,

$$H(\omega) = |H(\omega)|e^{j\varphi(\omega)} \tag{1}$$

where $|H(\omega)|$ is the magnitude response and $\varphi(\omega)$ is the phase response, which may be decomposed into the minimum-phase response $\mu(\omega)$ and the excess-phase response $\eta(\omega)$,

$$\varphi(\omega) = \mu(\omega) + \eta(\omega). \tag{2}$$

The group delay $\tau(\omega)$ of $H(\omega)$ is defined as the negative derivative of the phase response,

$$\tau(\omega) = -\mathrm{Im}\left\{\frac{d}{d\omega}\log H(\omega)\right\} = -\frac{d}{d\omega}\varphi(\omega). \tag{3}$$

In particular, the excess group delay $\gamma(\omega)$ is

$$\gamma(\omega) = -\frac{d}{d\omega}\eta(\omega). \tag{4}$$

Since the minimum-phase transfer function is

$$H_{\mathrm{mp}}(\omega) = |H(\omega)|e^{j\mu(\omega)}. \tag{5}$$

The HRTF may be written as the product of the minimum-phase transfer function and the excess-phase response,

$$H(\omega) = H_{\mathrm{mp}}(\omega)e^{j\eta(\omega)}. \tag{6}$$

Note that in the standard HRTF representation used in binaural synthesis, the excess-phase response $\eta(\omega)$ is a linear-phase characteristic, $\eta(\omega) = \omega\tau$.

2.1. Retained Coherence

The cross-coherence $\psi_{xy}(l)$ between signals $x(n)$ and $y(n)$ is

$$\psi_{xy}(l) = \frac{\sum_{n=0}^{\infty} x(n-l)y(n)}{\left[\sum_{n=0}^{\infty} x^2(n)\cdot\sum_{n=0}^{\infty} y^2(n)\right]^{\frac{1}{2}}}, \tag{7}$$

where the lag l is the time shift between the signals at which the coherence is evaluated. Note that the cross-choerence between signals $x(n)$ and $y(n)$ is their cross-correlation normalized by the geometric mean of the signal energies. The maximum of the cross-coherence over lag l,

$$\psi^* = \max_l \{\psi_{xy}(l)\}, \tag{8}$$

is an indication of the similarity between the signals $x(n)$ and $y(n)$.

Here we use the maximum of cross-choerence between an HRTF and its minimum-phase version as a means of quantifying the HRTF's minimum-phase content. This coherence maximum is called here the *retained coherence*, denoted by $\rho(\theta, \phi)$,

$$\rho(\theta, \phi) = \max_l \left\{ \frac{\sum_{n=0}^{\infty} h_{\mathrm{mp}}(n-l)\, h(n)}{\sum_{n=0}^{\infty} h^2(n)} \right\} \tag{9}$$

where $h(n)$ is the HRIR evaluated at azimuth θ and elevation ϕ, and $h_{\mathrm{mp}}(n)$ is its minimum-phase reconstruction.

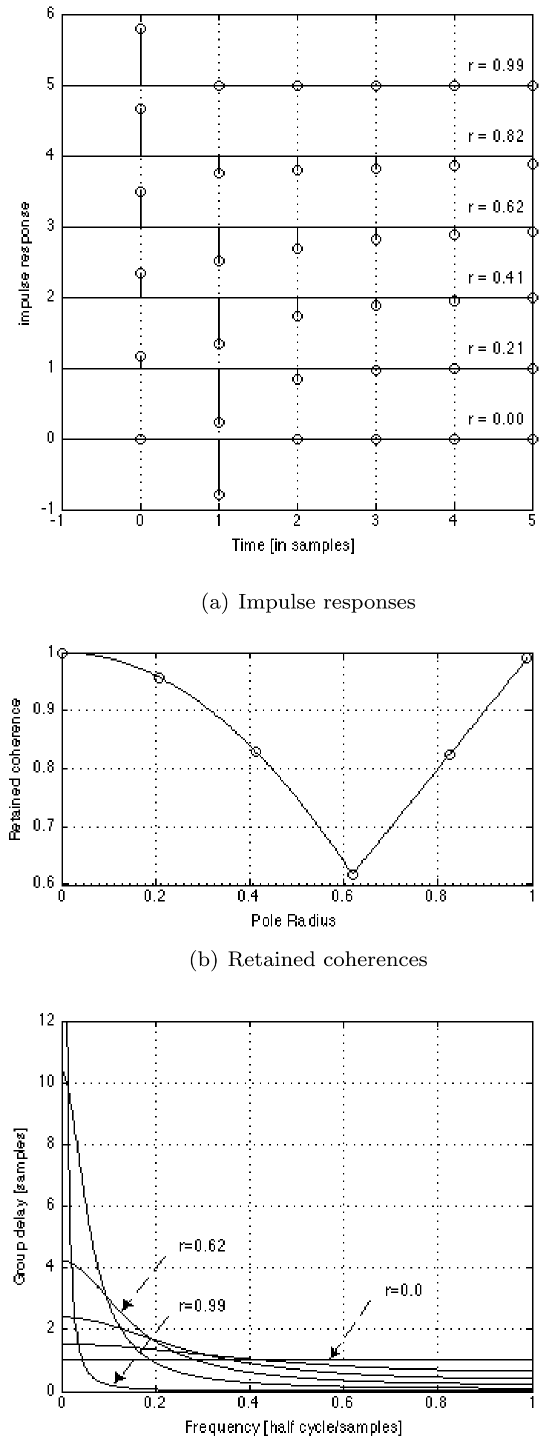
Note that since the cross-correlation is computed in discrete time, its maximum value can be more precisely estimated by quadratically interpolating the maximum [9]. If $h(n)$ consists of a pure delay and a minimum-phase sequence, the retained coherence $\rho(\theta, \phi)$ will be 1.0. Consequently, when the retained coherence is close to 1.0, $h(n)$ will be nearly identical in shape to $h_{\mathrm{mp}}(n)$.

To understand the properties of the retained coherence in the presence of linear-phase and non-pure-delay, consider a first-order allpass filter $G(z)$ with a simple pole at $z = r$ and a zero $z = 1/r$,

$$G(z) = \frac{-r + z^{-1}}{1 - rz^{-1}}. \tag{10}$$

Fig. 2(a) shows impulse responses associated with $G(z)$ for different pole radii r. At the extreme cases ($r = 0$ and $r = 0.99$), the impulse response is very nearly an impulse, and the retained coherences are very close to 1.0. On the other hand, impulse responses associated with immediate radii tend to be dispersed over time, reducing the retained coherences. For the first-order allpass filter $G(z)$, the retained coherence is given by

$$\rho = \max(|r|, 1 - r^2), \tag{11}$$



(a) Impulse responses

(b) Retained coherences

(c) Group delays

Fig. 2: The first-order allpass filter

and the associated group delay is

$$\tau(\omega) = \frac{1 - r^2}{1 - 2r\cos\omega + r^2}. \quad (12)$$

The retained coherence is shown in Fig. 2(b), and the associated group delay is shown in Fig. 2(c). At $r = 0$, the group delay is a constant function, clearly having a retained coherence of 1.0. At $r = 0.99$, even though the group delay has a sharp peak, it is flat for most of the band, and the retained coherence is almost 1.0. On the other hand, moderate peaks in the group delay such as at $r = 0.62$ correspond to a low retained coherence, indicating that such non-constant group delay features reduce retained coherence.

2.2. HRTF Retained Coherence

Three sets of HRTF data were used to explore HRTF phase. One set, "CCRMA HRTFs", consists of eight subjects we measured in the CCRMA recording studio using blocked meatus microphones as described in [10]. Swept sinusoids similar to those described in [11] were employed, with HRTFs measured every 15 degrees in azimuth and every 10 degrees in elevation, ranging from −40 degrees to +40 degrees. The second set, "Snapshot HRTFs", consists of 23 subjects measured using blocked meatus microphones. Impulse responses were measured every 30 degrees in azimuth and every 18 degrees in elevation, ranging from −36 degrees to +54 degrees, using Golay codes to measure the raw HRIRs. The third data set, "WK HRTFs" by Wightman and Kistler at the University of Wisconsin-Madison, consists of five subjects measured at 505 directions using open meatus microphones.

Example retained coherences for a single WK subject and a single CCRMA subject appear in Fig. 5 and 6. Note that most of the HRTF retained coherences are above 0.9, and that the HRTFs retained coherences below 0.9 are concentrated in the frontal and ipsilateral directions.

Fig. 3 and 4 show histograms of the retained coherences for all measured subjects in the WK and CCRMA data sets. As in the examples shown in Fig. 5 and 6, the vast majority of the retained coherences are above 0.9, while the open meatus WK HRTFs result in slightly larger retained coherences. Note that the distributions of retained coherence over azimuth and elevation exhibit similar patterns between the WK and CCRMA data sets, with the frontal and ipsilateral retained coherence being somewhat lower than at other directions. It should be noted that the Snapshot HRTF retained coherence histograms, not shown here, show a similar pattern with respect to arrival direction: they have a similar but slightly lower average retained coherence compared with that of the blocked meatus CCRMA HRTF set.

3. HRTF GROUP DELAY

3.1. Retained Coherence Interpretation

The retained coherence may be expressed as a weighted sum of phase errors:

$$\rho = \max_l \left[\frac{\mathcal{F}^{-1}\left\{H(\omega)H^*_{\mathrm{mp}}(\omega)\right\}}{\sum_\omega |H(\omega)|^2} \right] \quad (13)$$

$$= \max_l \left[\sum_\omega V(\omega)\,\cos(\omega l - \eta(\omega)) \right], \quad (14)$$

where the weighting $V(\omega)$ is proportional to the HRTF power at frequency ω,

$$V(\omega) = \frac{|H(\omega)|^2}{\sum_\omega |H(\omega)|^2}. \quad (15)$$

Note that the excess-phase is a linear function of frequency $\eta(\omega) = \omega l$, the summation in (14) is maximized by setting $l = \tau$, and a retained coherence of 1.0 is obtained. In general, the retained coherence is reduced from 1.0 in cases where the excess phase deviates from a linear characteristic, and the HRTF has energy at frequencies over which the deviations occurs. This leads the idea that the similarity between HRTFs and their minimum-phase counterparts can be found by examining the excess group delay and the associated magnitude response.

3.2. HRTF excess group delay

HRTF excess group delays were computed for the measured HRTFs. Many had non-constant excess group delays at high frequencies, particularly those at contralateral angles. Fig. 7 shows two typical excess group delays and magnitude responses, one at an ipsilateral angle and the other at a contralateral angle. The excess group delay in Fig. 7(a) is characterized by a large narrow-bandwidth peak. There is

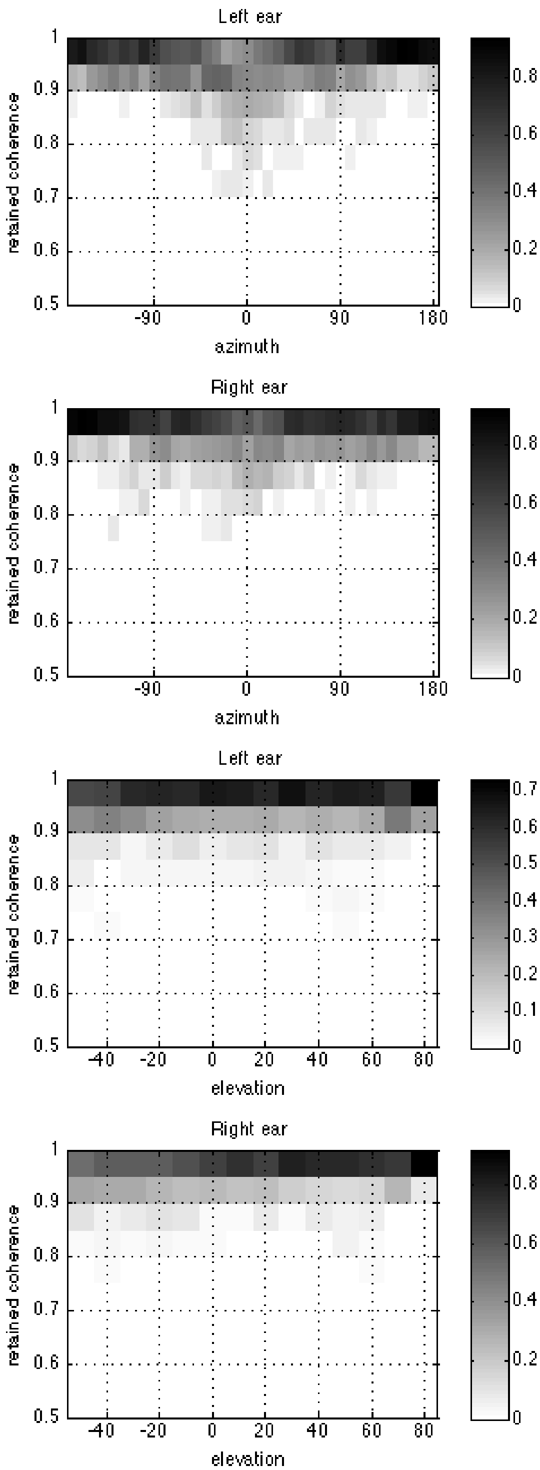


Fig. 3: Distribution of retained coherence for HRTFs measured with open meatus microphones

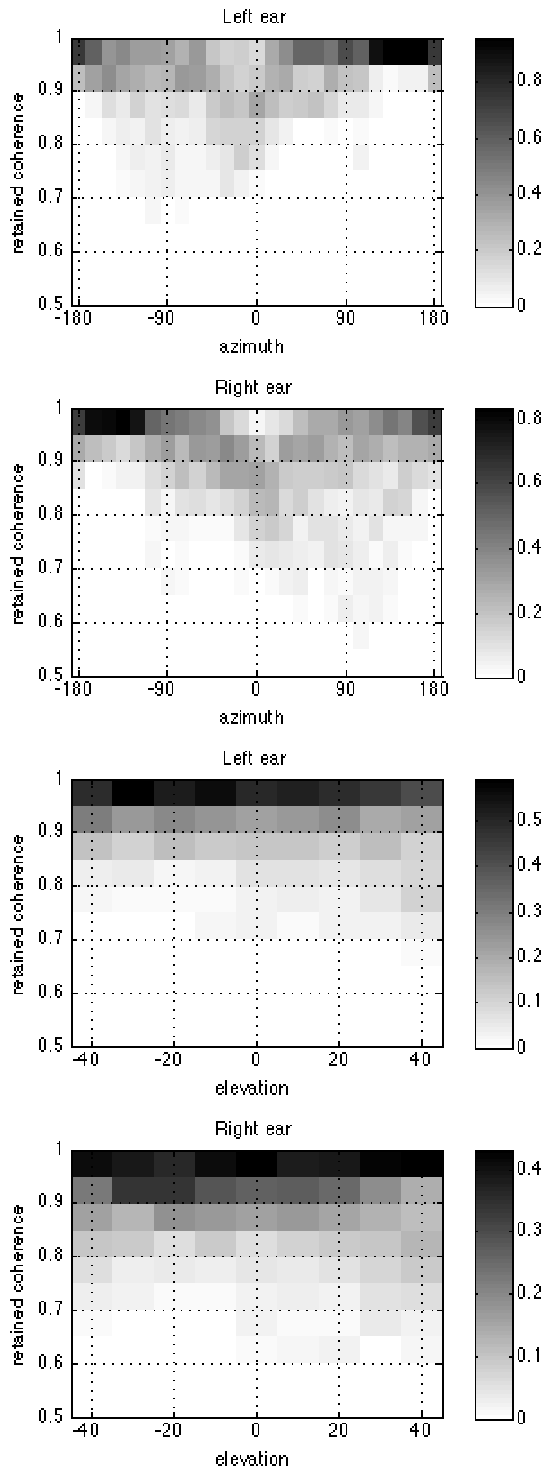


Fig. 4: Distribution of retained coherence for HRTFs measured with blocked meatus microphones

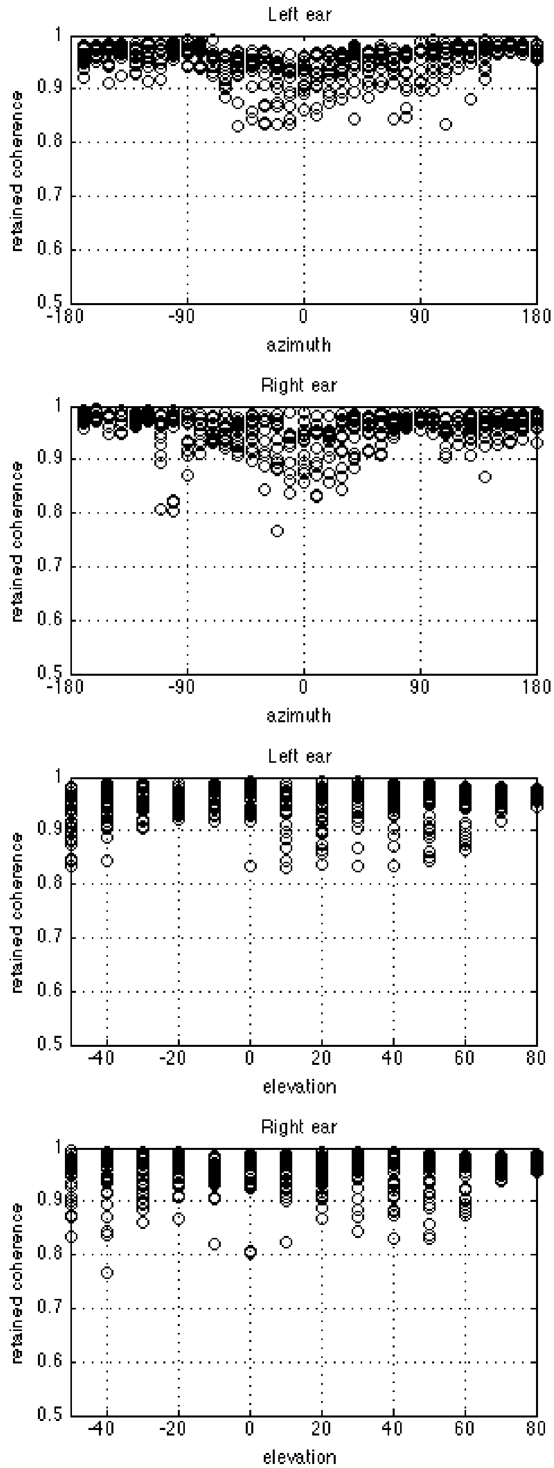


Fig. 5: Retained coherence for a single subject: HRTF measured with open meatus microphones

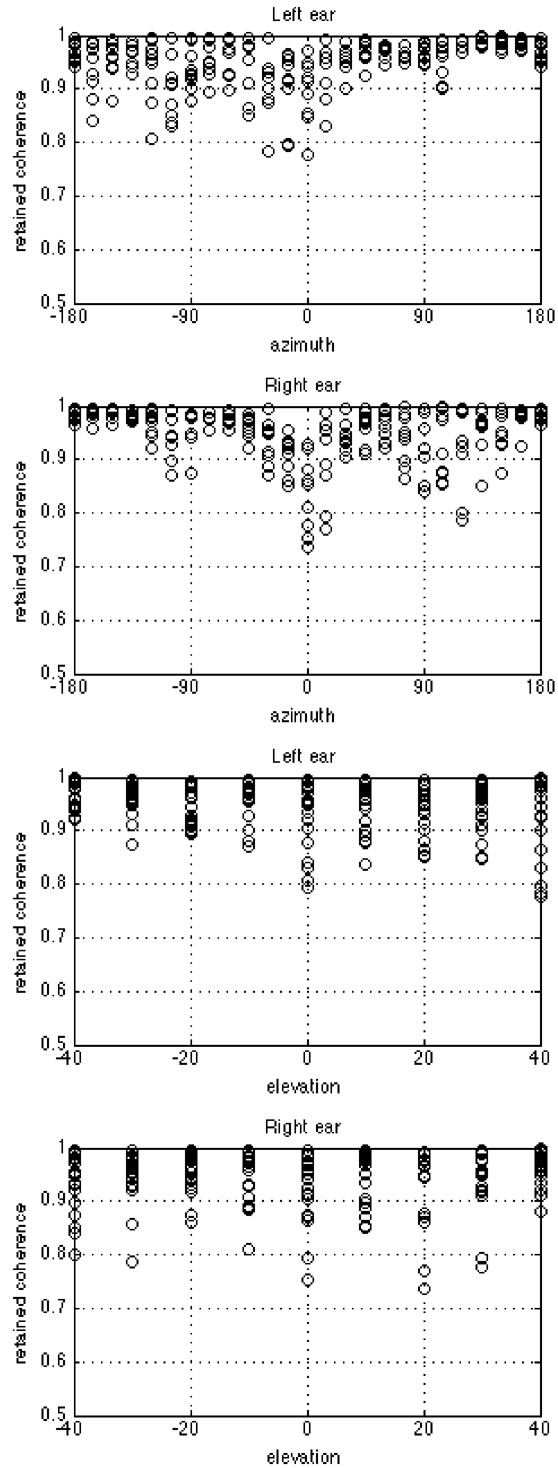


Fig. 6: Retained coherence for a single subject: HRTF measured with blocked meatus microphones

a spectral notch at the same frequency, and therefore the retained coherence remains close to 1.0. This narrow-band feature likely results from an isolated non-minimum-phase zero.

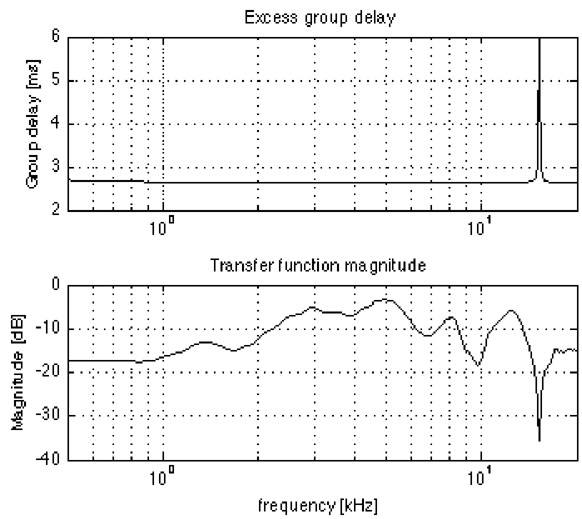
Similar regions of non-constant excess group delay are found in Fig. 7(b). The sharp group delay peaks correspond to spectral notches, such as the notch labeled "B" in Fig. 7(b), and therefore have little effect on the retained coherence. Note that the notch labeled "C" corresponds to a region of constant group delay, and, therefore, to a minimum-phase zero. Another feature type occasionally present is a broad group delay peak corresponding to a region of moderate spectral energy, such as the feature labeled "A" in Fig. 7(b). In this case, the retained coherence is reduced from 1.0.

4. SUMMARY

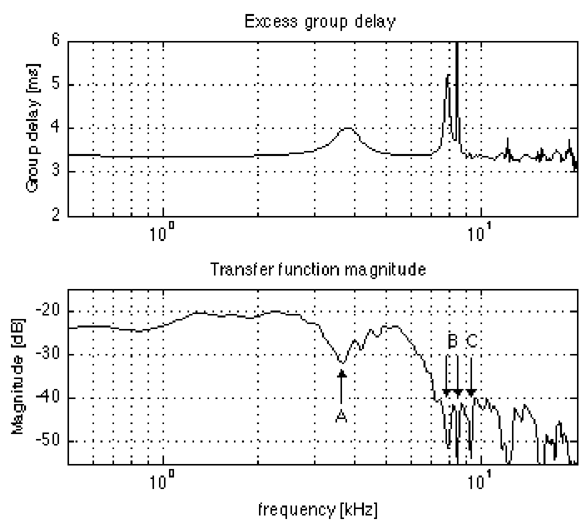
Retained coherence—the maximum of the cross-coherence between an HRIR and its minimum-phase version—was defined and used to evaluate the minimum-phase content of HRTFs. HRTFs were seen to be well-modeled by pure delays followed by minimum-phase filters: a substantial majority of HRTFs had retained coherences greater than 0.9, and rarely less than 0.8. A small difference was noted between HRTF measurements made with blocked meatus and open meatus microphones, with the open meatus measurements producing somewhat greater retained coherences. HRTFs having non-minimum-phase content were found in frontal and ipsilateral directions. These HRTFs had non-minimum-phase zeros at high frequencies rarely below 8 kHz.

5. ACKNOWLEDGEMENT

We would like to thank CCRMA students and staff for their assistance and participation in the HRTF measurements.



(a) Retained coherence : 0.977



(b) Retained coherence : 0.791

Fig. 7: Excess group delay and magnitude response at (a) azimuth=90°, elevation=0° (b) azimuth=255°, elevation=0°

6. REFERENCES

[1] D. J. Kistler, F. L. Wightman, "A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction," J. Acoust. Sot. Am., Vol. 91, pp. 1637-1647, 1992.

[2] J.-M. Jot, V. Larcher, O. Warusfel, “Digital signal processing issues in the context of binaural and transaural stereophony,” presented at the 98th Audio Engineering Society Convention, February, 1995.

[3] J. Huopaniemi, M. Karjalainen, “Review of Digital Filter Design and Implementation Methods for 3-D Sound,” presented at the 102th Audio Engineering Society Convention, February, 1997.

[4] V. Larcher, J.-M. Jot, G. Guyard, and O. Warusfel, “Study and Comparison of Efficient Methods for 3-D Audio Spatialization Based on Linear Decomposition of HRTF Data,” presented at the 108th Audio Engineering Society Convention, February, 2000.

[5] J. Plogsties, P. J. Minnaar, S. K. Olesen, F. Christensen, H. Moller, “Audibility of allpass components in head-related transfer functions,” presented at the 108th Convention of the Audio Engineering Society, Paris, France, February 2000.

[6] S. Mehrgardt, V. Mellert, “Transformation characteristics o the external human ear,” Journal of the Acoustical Society of America, Vol. 61, pp. 1567-1576, 1977.

[7] A. Kulkarni, S. K. Isabelle, H. S. Colburn, “On the minimum-phase approximation of head-related transfer functions,” IEEE ASSP Workshop on Application of Signal Processing to Audio and Acoustics, 1995, pp. 84-87.

[8] A. Kulkarni, S. K. Isabelle, H. S. Colburn, “Sensitivity of human subjects to head-related transfer-function phase spectra,” J. Acoust. Soc. Am. 105 (5), May 1999.

[9] J. O. Smith, “Spectral audio signal processing”, March 2007 Draft, `//ccrma.stanford.edu/~jos/sasp/`, online book, accessed July 2008.

[10] J. S. Abel, S. H. Foster, “Measuring HRTFs in a Reflective Environment,” International Community for Auditory Display, November 1994.

[11] A. Farina, “Simultaneous Measurement of Impulse Response and Distortion with a Swept-Sine Technique,” presented at the 108th Convention of the Audio Engineering Society, Paris, France, February 2000.