DAFx'04 Invited paper

BINAURAL SOURCE LOCALIZATION

Harald Viste

Audiovisual Communications Lab
Swiss Federal Institute of Technology Lausanne
Switzerland
harald.viste@epfl.ch

Gianpaolo Evangelista

Dept. of Physical Sciences
Federico II University of Naples
Italy
gianpaolo.evangelista@na.infn.it

ABSTRACT

In binaural signals, interaural time differences (ITDs) and interaural level differences (ILDs) are two of the most important cues for the estimation of source azimuths, i.e. the localization of sources in the horizontal plane. For narrow band signals, according to the duplex theory, ITD is dominant at low frequencies and ILD is dominant at higher frequencies.

Based on the STFT spectra of binaural signals, a method is proposed for the combined evaluation of ITD and ILD for each individual spectral coefficient. ITD and ILD are related to the azimuth through lookup models. Azimuth estimates based on ITD are more accurate but ambiguous at higher frequencies due to phase wrapping. The less accurate but unambiguous azimuth estimates based on ILDs are used in order to select the closest candidate azimuth estimates based on ITDs, effectively improving the azimuth estimation. The method corresponds well with the duplex theory and also handles the transition from low to high frequencies gracefully.

The relations between the ITD and ILD and the azimuth are computed from a measured set of head related transfer functions (HRTFs), yielding azimuth lookup models. Based on a study of these models for different subjects, parametric azimuth lookup models are proposed. The parameters of these models can be optimized for an individual subject whose HRTFs have been measured. In addition, subject independent lookup models are proposed, parametrized only by the distance between the ears, effectively enabling source localization for subjects whose HRTFs have not been measured.

1. INTRODUCTION

Binaural source localization is the problem of estimating the location of a sound source based on the signals observed at the entrances of the two ears. For localization of sources in the horizontal plane, i.e. the estimation of azimuth angles, the differences between the two ear signals are most important. These are described by the relation between the HRTFs for the two ears. In this paper we consider the problem of estimating source azimuth angles.

1.1. Background

Over the past decades, several computational models for the estimation of source azimuths in binaural signals have been proposed. Many of these [1, 2] are based on the coincidence model proposed by Jeffress in 1948. This is a model of the neural system where nerve impulses from each of the two ears propagate along delay lines in opposite directions. At the position along the delay lines where the impulses coincide a nerve cell is excited, effectively transforming time information into spatial information. This model corresponds to evaluating ITDs by means of a running short-time cross-correlation function between the ear input signals. Based on Jeffress model, several extensions have been proposed that take into account ILDs, such as the models by Lindemann [3] and Gaik [4]. An overview of these and other models of binaural perception is given in [5]. Most of these models work by decomposing the ear input signals into perceptual bands and estimate the interaural cues in these bands. When several sources at different locations have significant energy within a given perceptual band, the resulting azimuth estimates for that band will not, in general, correspond to any of the actual azimuths of the sources. In some cases, therefore, it can be advantageous not to be limited by the frequency resolution of the human auditory system, but rather to estimate the azimuths in individual narrow frequency bands.

1.2. Contribution

When HRTFs have been measured at several different azimuth angles for each of the two ears, the differences between these two sets of HRTFs describe the ILDs and ITDs as functions of azimuth (and frequency). This means that, in an observed signal, an ILD estimate can be compared with the HRTF data sets in order to obtain an estimate of the source azimuth. This is referred to as **HRTF data lookup**, yielding azimuth estimates based on ILD only. Similarly, the ITDs can be used for HRTF data lookup of azimuths based on ITD only.

In this paper we propose a method for the estimation of source azimuths through the joint evaluation of ILDs and ITDs. The method is based on the short-time Fourier transform (STFT) spectra of the input signals, and the ILD and ITD is estimated for each spectral coefficient. On one hand, the azimuth estimates based on ILDs have a relatively large standard deviation. On the other hand, the azimuth estimates based on ITD have smaller standard deviation, but are ambiguous. By jointly evaluating these quantities the ILDs are used in order to resolve the ITD ambiguities, effectively improving the azimuth estimates. Since the method is based on the STFT it is computationally fast and has a simple reconstruction scheme that is highly useful, e.g. in source separation applications.

We also propose a parametric model for the relation between azimuth angle and interaural cues (ILD and ITD). In this model the parameters are optimized with respect to an individual head for which the HRTFs have been measured. Similarly to the azimuth estimation by HRTF data lookup, it is possible to lookup the azimuths by use of this model. This is referred to as **individual**

model lookup. This method has the advantage that the azimuth lookup is faster. However, the azimuth estimates are not as accurate as those obtained by HRTF data lookup. In addition, it still requires the HRTFs to be measured for the individual head in order to determine the parameters.

Based on the study of the parameters of the individual model for each of the 45 subjects in the CIPIC database of HRTFs [6] we propose a generic model for the relation between azimuth angle and ILDs and ITDs that only depends on one parameter, namely the distance between the ears. This model can be used with any head and does not require the measurement of HRTFs. Estimation of azimuths by **average model lookup** gives results comparable to those obtained by inidividual model lookup.

In Section 2 the estimation of ILDs and ITDs for individual spectral coefficients in the STFT spectra is discussed. In Section 3 we propose a method for joint evaluation of these cues. The parametric model for ILD and ITD is presented in Section 4. In Section 5 the parameters of the individual model are studied and the average model is proposed. The application to source localization is studied in Section 6. In Section 7 we draw the conclusions.

2. CUE ESTIMATION

From the two observed ear entrance signals, the STFT spectra are computed. These are denoted by $X_{\text{left}}(k,q)$ and $X_{\text{right}}(k,q)$, where k and q are the time and frequency indexes, respectively. In this time-frequency representation, the spatial cues can be easily estimated for each spectral coefficient (left/right pair) individually. The ILDs in dB are given by

$$\Delta L(k,q) = 20\log_{10}\left|\frac{X_{\text{right}}(k,q)}{X_{\text{left}}(k,q)}\right|. \qquad (1)$$

This is simply the ratio, measured in dB, of the STFT magnitudes of the right and left ear signals. Similarly, the ITDs are estimated as

$$\Delta T_p(k,q) = \frac{-\Delta P_p(k,q)}{q}\frac{L}{2\pi}, \qquad (2)$$

where L is the window length in the STFT. The interaural phase differences $P_p(k,q)$ are given by

$$\Delta P_p(k,q) = \arg\frac{X_{\text{right}}(k,q)}{X_{\text{left}}(k,q)} + 2\pi p. \qquad (3)$$

Each spectral coefficient represents a periodic and narrow band signal. The phase of a periodic signal can only be estimated up to an integer multiple of 2π. This is reflected by the integer parameter p in the estimates of ITD and interaural phase difference. The practical significance of this is that, for a given frequency, several different source locations yield the same phase difference between the two ear input signals. This is equivalent to the spatial aliasing seen in beamforming techniques. The parameter p indexes these positions, with $p = 0$ corresponding to the source position closest to zero azimuth, $\theta = 0$. A negative value of p corresponds to a position on the left side (negative θ). Positive p corresponds to positions on the right side. In this case, possible values of p depend on the physical layout of the sensors and sources. The frequency whose period equals twice the largest possible delay between the two ears corresponds to the highest frequency for which the phase can be estimated without ambiguity. Below this frequency only $p = 0$ is physically realizable. For an average head size the phase ambiguities occur for frequencies above approximately 1500 Hz.

3. ESTIMATION OF AZIMUTH ANGLES

In order to relate the ILDs to the ITDs, a common reference frame is needed such as the azimuth angle. Using measured HRTFs, the azimuth can be estimated from ILDs and ITDs by HRTF data lookup.

3.1. HRTF data lookup

Based on the HRTFs measured at different azimuth angles, the ILD and ITD can be described as function of azimuth and frequency. Since the HRTFs are assumed to be time-invariant, there is no dependency on the time index k. Instead, the HRTFs depend on the azimuth angle θ. By changing the role of the time index k with that of the azimuth angle θ and by using the left and right HRTFs as functions of azimuth and frequency, $H_{\text{right}}(\theta,q)$ and $H_{\text{left}}(\theta,q)$, as the signal spectra in Equations (1) and (2), we obtain the HRTF data lookup models for level difference, $\Delta L(\theta,q)$, and time difference, $\Delta T(\theta,q)$, as functions of azimuth angle θ and frequency index q. In the computation of the ITD lookup model special care must be taken to "unwrap" the phase, i.e. to determine the correct choice of p for all frequencies and azimuths.

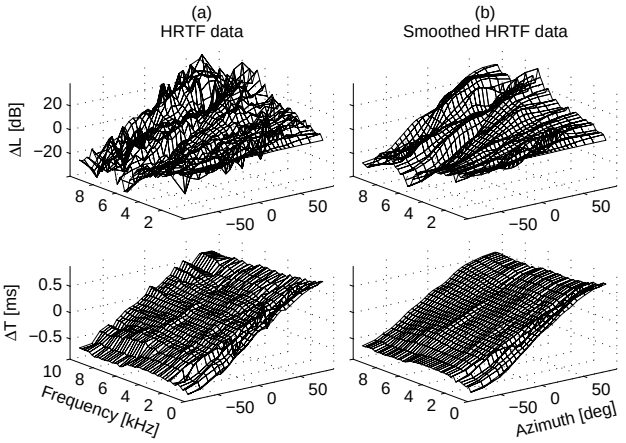

Figure 1: *Interaural level and time differences as functions of azimuth angle and frequency. (a): HRTF data lookup. (b): HRTF data lookup smoothed across azimuth.*

The ILD and ITD as functions of azimuth and frequency for one particular head are shown in Figure 1. The panels in the left column show the ILD and ITD as computed from the measured HRTFs. In the right column, the same functions are shown after smoothing in the azimuth direction. No processing across frequency was performed.

When a sound source is located on one side of the head, the source signal will arrive first at the ear on the same side. Also, the signal level will normally be stronger at the ear facing the source than at the ear on the opposite side. Intuitively, the farther to the side a source is located, the larger the level and time differences between the ears should be. The smoothed HRTF data confirm this intuitive facts. Both functions are relatively monotonic in azimuth.

Based on the ILD estimate (1) for a given left/right pair of spectral coefficients, $X_{\text{left}}(k,q)$ and $X_{\text{right}}(k,q)$, the azimuth angle can be looked up in the smoothed HRTF data $\Delta L(\theta,q)$. This yields azimuth estimates based on ILD only, denoted $\theta_L(k,q)$.

Similarly, each ITD estimate (2) can be looked up in $\Delta T(\theta, q)$. Due to the phase ambiguity, this results in multiple possible azimuth estimates, denoted $\theta_{T_p}(k, q)$, indexed by p.

The ITD is usually a relatively smooth function of azimuth, as seen in Figure 1. This means that the standard deviation of the azimuth estimates based on ITD is relatively small. However, there may be several possible azimuth candidates due to the phase ambiguity. The ILD is a more complex function of azimuth and must be smoothed across azimuth in order to become useful for azimuth lookup. Consequently, the azimuth estimates based on ILDs have a much larger standard deviation than those based on ITD. In addition, the ILDs as function of azimuth are not, in general, monotonic for all frequencies. When this is the case, the azimuth lookup is non-unique, yielding multiple possible azimuth estimates. For the examples in this paper the azimuth estimates closest to 0 degrees were chosen whenever this was the case.

3.2. Joint evaluation of ILD and ITD

Since both, the ILD and the ITD, are related to the azimuth, they can also be related to each other. We propose a method for the joint evaluation of these quantities in order to improve the azimuth estimates. Briefly explained, the noisy $\theta_L(k, q)$ provides a rough estimate of the azimuth for each left/right spectral coefficient pair. Then, this estimate is refined by choosing the $\theta_{T_p}(k, q)$ that lies closest. The combined azimuth estimate is then given by

$$\theta(k, q) = \theta_{T_i}(k, q)\Big|_{i=\arg\min_p\left(\left|\theta_L(k,q)-\theta_{T_p}(k,q)\right|\right)} . \quad (4)$$

Effectively, the ILD estimate is used in order to choose the "correct" parameter p in the ITD estimate. The azimuth estimate based on ITD is chosen since this estimate is "more precise", i.e. the standard deviation of these estimates are smaller. The general processing is illustrated in Figure 2.

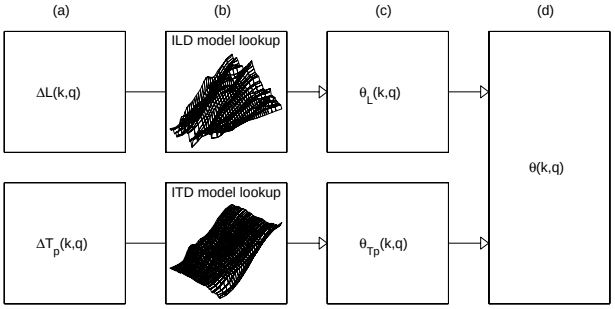

Figure 2: *Combined evaluation of ILDs and ITDs for the estimation of azimuths. (a): The interaural cues are estimated by use of the STFT. (b): The relation between these cues and azimuth angle is described by the ILD and ITD in the smoothed HRTF data. (c): Azimuth estimates are found by lookup of the interaural cues in the HRTF data models. (d): Final azimuth estimates are obtained by combined evaluation of the azimuths estimated from the ILDs and ITDs.*

Figure 3 shows some experimental results to further illustrate the processing. In order to assess the performance of the proposed method for different frequencies, a white noise signal was chosen as source signal. This signal was then filtered with HRTFs for five different heads, each using a different azimuth angle for the

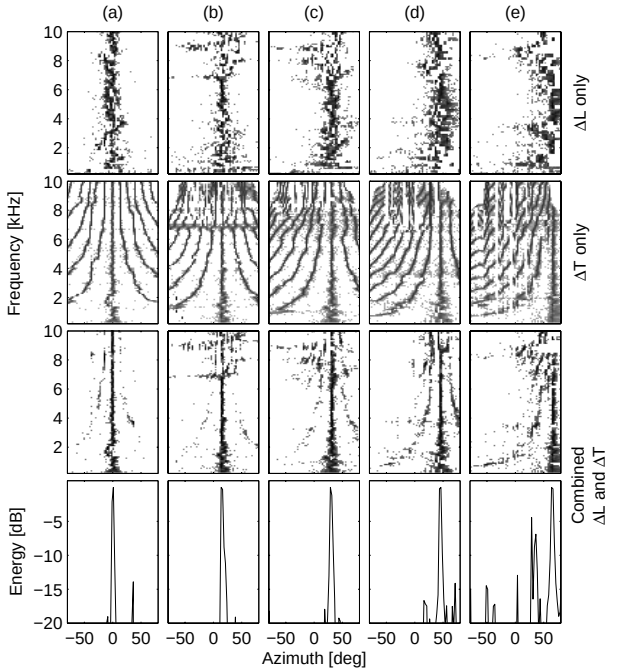

Figure 3: *Histograms of azimuth estimates for 5 different heads and azimuth angles, at 0, 15, 30, 45 and 65 degrees, (a)–(e), respectively. First row: based on ILD only. Second row: based on ITD only. Third row: based on our method for combined evaluation of ILD and ITD. Bottom row: marginal histograms for our method.*

source position. A window length of about 10 ms was chosen in the computation of the STFT spectra, and the interaural cues were estimated by means of (1) and (2). The five different situations are shown in the different columns in the figure, with azimuth angles of 0, 15, 30, 45 and 65 degrees (on the right side of the head), respectively. The panels in the first row show two-dimensional histograms as function of azimuth and frequency, based on azimuths estimated from ILD only. These histograms imply, as mentioned in Section 3.1, that the azimuth estimates based on ILD have a larger standard deviation than those based on ITD. The precision decreases with increasing azimuth. In addition, at low frequencies, the ILDs are small and are virtually useless for the estimation of the azimuth. The panels in the second row of Figure 3 show similar histograms based on azimuths estimated from ITDs. Above approximately 1-2 kHz, these azimuth estimates are ambiguous. For a given frequency this is seen as several equally strong peaks at different azimuths. These correspond to different choices of p in (2). As the frequency increases, more values of p are possible. Additionally, the distance between the peaks decreases with increasing frequency, making the ITD estimates less useful at higher frequencies. In the third row the results that were obtained by applying the proposed method are shown. Visually explained, the estimates based on ILDs (first row), are used in order to select the right p in the estimates based on ITDs (second row). Note that this is done independently for each spectral coefficient in the STFT spectra: no processing is performed across frequency.

The last row of panels shows the one-dimensional marginal

histograms as function of azimuth (i.e. summing over all frequencies), based on the azimuths computed by the proposed method. The strongest peaks are found at the true azimuth angles of 0, 15, 30, 45 and 65 degrees, respectively.

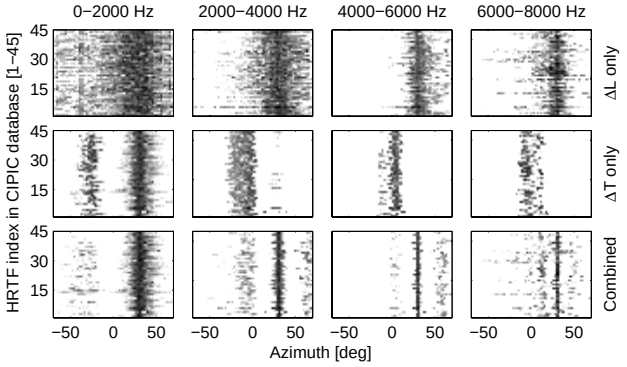

Figure 4: *Energy weighted histograms of azimuth estimates for 45 subjects in 4 different frequeency bands; using HRTF data lookup.*

The example with a source located at 30 degrees azimuth, as shown in Figure 3(c), was repeated for all the 45 heads in the CIPIC database. For each head, the one-dimensional marginal histograms, as function of azimuth, were computed in four different frequency ranges, namely 0–2 kHz, 2–4 kHz, 4–6 kHz and 6–8 kHz. The results are shown in Figure 4. The four columns correspond to the different frequency ranges. The three rows correspond to the different azimuth estimates, i.e. based on ILD only (top), ITD only (middle), and the proposed method for combined ILD and ITD (bottom). Each panel shows the marginal histogram, as seen in the bottom row in Figure 4, for all the 45 different heads along the ordinate axis. For all the frequency ranges and for most of the heads the azimuth estimates have been improved by application of the proposed method. Most of the ambiguous azimuths based on ITD only have been resolved. In addition, the standard deviation is smaller than for the azimuth estimates based on ILD only.

4. PARAMETRIC ILD AND ITD MODELS

In the previous section, we described how to apply HRTF data in order to lookup azimuths. In this section we propose a parametric model for the relation between azimuth angles and ILDs and ITDs. This model allows a simpler lookup of azimuths, but at the cost of decreased accuracy. More importantly, it serves as the basis and motivation for the generic model that we propose in Section 5.

4.1. Interaural time differences

Based on simple geometric considerations, the following formula for the ITD was proposed in [7]:

$$\Delta T(\theta) = \frac{r(\sin\theta + \theta)}{c}, \tag{5}$$

where r is the "head radius", and c is the wave propagation speed. In reality, the ITD is slightly larger than this due to the fact that the head is not perfectly spherical. In addition, the time difference is also somewhat larger at low frequencies, as has been observed in [8]. In order to take this into account, we propose to use a frequency dependent scaling factor α_q,

$$\Delta T(\theta, q) = \alpha_q \frac{r(\sin\theta + \theta)}{c}. \tag{6}$$

4.2. Interaural level differences

As implied by the data shown in Figure 1, the ILD is a much more complex function of azimuth and frequency. Based on a study of the HRTFs in the CIPIC database, we propose the following model:

$$\Delta L(\theta, q) = \beta_q \frac{\sin\theta}{c}, \tag{7}$$

with frequency dependent scaling factor β_q. The effect of the source distance can be largely neglected, as indicated in [9, 10]. Based on the observation that the ILD is periodic in θ, a Fourier series expansion of the ILD was proposed in [11]. Our model is similar to this (single-term expansion), with the exception that we only consider the range $-90^\circ \leq \theta \leq 90^\circ$.

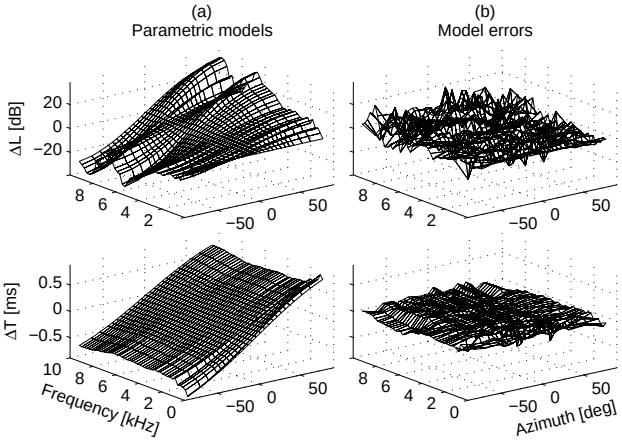

Figure 5: *Interaural level and time differences as functions of azimuth angle and frequency. (a): Individual parametric models. (b): Model errors.*

4.3. Frequency dependent scaling factors

The ILD and ITD models in (7) and (6) can be optimized for a given head by finding the scaling factors α_q and β_q that give the closest match to the smoothed HRTF data. For the head whose HRTF data is shown in Figure 1, the best matching parametric models were found. The individual parametric models and the model errors are shown in Figure 5.

For the estimation of azimuth based on ILDs and ITDs, (Figure 2(b)), the HRTF data can be replaced by the parametric model. The experiment shown in Figure 4 was repeated, but with the use of parametric models. For each of the 45 heads, the best matching parametric model was found and used for azimuth lookup. The results are shown in Figure 6. In this case, the azimuth estimates are not as accurate as when the HRTF data were used for azimuth lookup. In particular, the estimation errors are significant at higher frequencies. The estimation error is mainly due to the model error in the ILD model, as shown in Figure 5. For most heads, however,

the model is useful up to about 6 kHz. In any case, the use of the proposed method for the joint evaluation of ILD and ITD yields sharper peaks in the azimuth histograms.

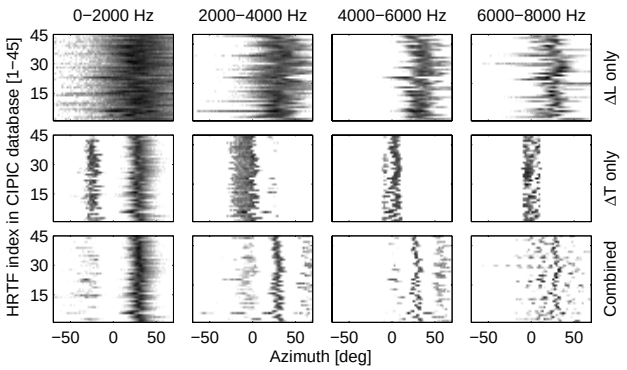

Figure 6: *Energy weighted histograms of azimuth estimates for 45 subjects in 4 different frequeency bands; using individual model lookup.*

5. AVERAGE ILD AND ITD MODELS

In order to obtain ILD and ITD models for a given head HRTFs must be measured for a range of different azimuth angles. This can be a tedious task. Also the parametric models that were proposed rely on the HRTF data for estimating the frequency dependent scaling factors. In this respect, the parametric models do not present an advantage over the HRTF data. In fact, the parametric models are less accurate.

The scaling factors for the ILD model, β_q, and ITD model, α_q, were computed as functions of frequency for the 45 subjects in the CIPIC database. These are shown in gray in Figure 7. Qualitatively, all these quantities follow the same trend, at least up to about 7 kHz. Above this frequency, the ILD scaling factors vary highly among the different heads.

The black lines in Figure 7 show the scaling factors averaged over all heads. If the average scaling factors are used in the parametric ILD and ITD models, these models only depend on one parameter, namely the head radius r. This can easily be measured, and consequently these average parameter models can be employed for any head without the need to measuring the specific HRTFs.

In order to compare the accuracy of the individual parametric models and the average models, the model errors were computed for each head (relative to the smoothed HRTF data). The absolute value of these errors were then averaged over all azimuths and heads. Figure 8 shows these errors as functions of frequency. The ITD models are shown in the top row, and the ILD models in the bottom row. Columns (a) and (b) correspond to the individual parametric models and the average parameter models, respectively. The average models (b) are almost as good as the individual models (a), and only a slight increase in error can be observed. However, above about 6-8 kHz the accuracy of all the models is significantly worse than that attained by the smoothed HRTF data.

The performance of the average models for estimation of azimuths is shown in Figure 9. The accuracy of azimuth estimates

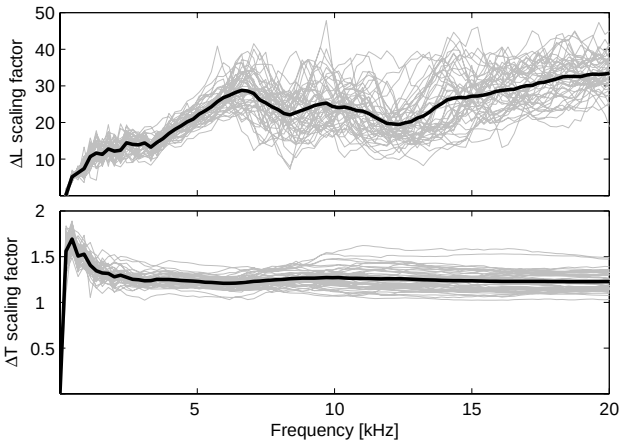

Figure 7: *Frequency-dependent scaling factors. Top: ILD model scaling factor, β_q. Bottom: ITD model scaling factor, α_q. The factors optimized for each of the 45 heads are shown in gray, and the average over these is shown in black.*

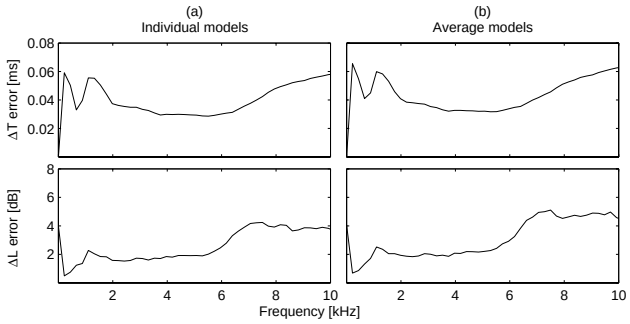

Figure 8: *Absolute model errors averaged over all azimuths and heads, as function of frequency. (a): Individual models. (b): Average parameter models.*

based on the average models are comparable to those based on the individual parametric models, as shown in Figure 6.

6. SOURCE LOCALIZATION

So far, only a single source was considered and white noise was used as the source signal in order to study the accuracy of azimuth estimates at different frequencies. Since the proposed method for joint evaluation of ILDs and ITDs is based on interaural cues in narrow bands independently it is also applicable in situations where several sources are located at different locations.

In the following experiment, three different harmonic tones were chosen as source signals, i.e. three consecutive half-notes played by an alto trombone. The binaural signal was obtained by filtering these sources with the HRTFs at different azimuth angles. Figure 10 shows histograms of the estimated azimuth angles in this mixture. The histograms have been energy weighted, i.e. each azimuth estimate is weighted by the energy of the corresponding spectral coefficient. The three columns show the results obtained by use of HRTF data, individual models, and average models, respectively. The panels in the first row show the results

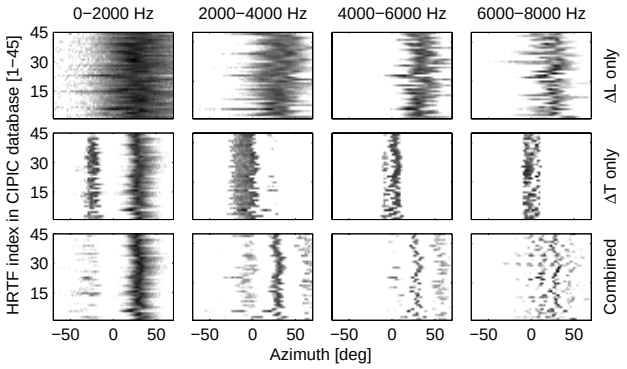

Figure 9: *Energy weighted histograms of azimuth estimates for 45 subjects in 4 different frequeency bands; using average model lookup.*

for azimuth estimates based on ILD only. Since the histograms are weighted, the different harmonics can be observed as strong peaks (dark horizontal lines). However, these are quite wide and do not provide very accurate estimates of the azimuth. In the second row, the estimates based on ITD are shown. The peaks corresponding to the harmonics are much narrower, but ambiguous above about 1500 Hz. In the third column the results obtained by the proposed method for joint evaluation of ILD and ITD are shown. For all three models, the different harmonics are well aligned about the true source azimuths of -30, 15 and 45 degrees, respectively. The marginal histograms are shown in the bottom row.

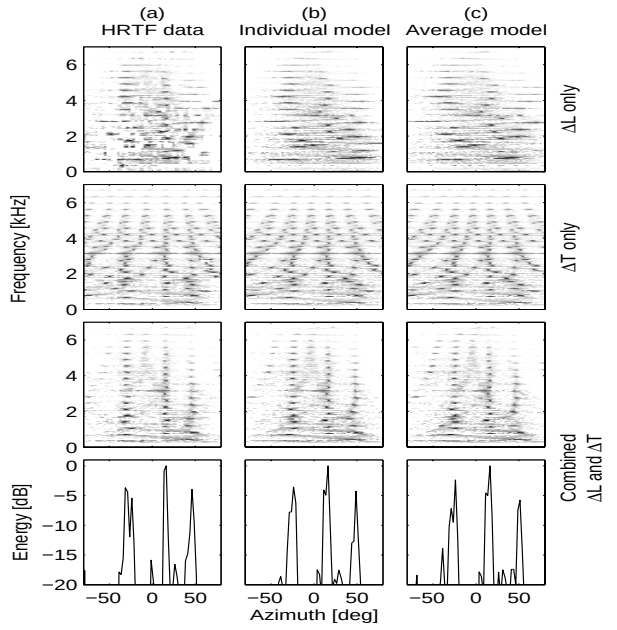

Figure 10: *Histograms of azimuth estimates in different frequency bands, comparing the three different models for azimuth lookup. (a): HRTF data. (b): Individual parametric model. (c): Average parameter model.*

7. CONCLUSIONS

We have presented a method for localization of source in the horizontal plane based on binaural signals. The method is narrow-band and short-time since it works with each spectral coefficient independently, effectively enabling the tracking of sources when the sources or sensors move [12]. We have also proposed a parametric head model and an average parameter model that can be used for a head whose HRTFs have not been measured.

8. REFERENCES

[1] Philip X. Joris, Philip H. Smith, and Tom C. T. Yin, "Coincidence detection in the auditory system: 50 years after jeffress," *Neuron*, vol. 21, pp. 1235–1238, December 1998.

[2] Jens Blauert, *Spatial Hearing*, MIT press, 2001.

[3] W. Lindemann, "Extension of a binaural cross-correlation model by contralateral inhibition. I. simulation of lateralization for stationary signals," *Journal of the Acoustical Society of America*, vol. 80, no. 6, pp. 1608–1622, December 1986.

[4] Werner Gaik, "Combined evaluation of interaural time and intensity differences: Psychoacoustic results and computer modeling," *Journal of the Acoustical Society of America*, vol. 94, no. 1, pp. 98–110, July 1993.

[5] Richard M. Stern and Constantine Trahiotis, *Models of Binaural Perception*, chapter 24, pp. 499–531, In Gilkey and Anderson [13], 1997.

[6] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, "The CIPIC HRTF database," in *Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, Mohonk, New York, USA, October 2001, pp. 99–102.

[7] R. S. Woodworth and H. Schlosberg, *Experimental psychology*, Holt, New York, 1954.

[8] Frederic L. Wightman and Doris J. Kistler, *Factors Affecting the Relative Salience of Sound Localization Cues*, chapter 1, pp. 1–23, In Gilkey and Anderson [13], 1997.

[9] Richard O. Duda and William L. Martens, "Range dependence of the response of a spherical head model," *Journal of the Acoustical Society of America*, vol. 104, no. 5, pp. 3048–3058, 1998.

[10] Richard O. Duda and William L. Martens, "Range-dependence of the HRTF for a spherical head," in *Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, Mohonk, New York, USA, October 1997.

[11] Richard O. Duda, *Elevation Dependence of the Interaural Transfer Function*, chapter 3, pp. 49–75, In Gilkey and Anderson [13], 1997.

[12] Harald Viste, *Binaural Localization and Separation Techniques*, Ph.D. thesis, Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland, 2004.

[13] Robert H. Gilkey and Timothy R. Anderson, Eds., *Binaural and Spatial Hearing in Real and Virtual Environments*, Lawrence Erlbaum Associates, 1997.