

An estimation method of interaural time differences from measured head-related impulse responses

Masashi Toyoda^{1,*}, Kanji Watanabe^{2,†}, Yukio Iwaya^{1,‡} and Yôiti Suzuki^{1,§}

¹Research Institute of Electrical Communication and Graduate School of Information Sciences, Tohoku University, 2-1-1 Katahira, Aoba-ku, Sendai, 980-8577 Japan

²Faculty of Engineering, University of Yamanashi, 4-3-11 Takeda, Kofu, 400-8511 Japan

(Received 5 December 2005, Accepted for publication 25 January 2006)

Keywords: HRIR, HRTF, ITD, Cross correlation function
PACS number: 43.60.Lq, 43.66.Jh [doi:10.1250/ast.27.239]

1. Introduction

Several methods of estimating the interaural time difference (ITD) from a measured set of head-related impulse responses (HRIR) have been proposed. Some examples are as follows:

- (1) The ITD is estimated as a difference between two initial delays calculated from the HRIRs of each ear [1].
- (2) The ITD is estimated as a lag of the time difference obtained by the maximum of the cross correlation coefficient between the HRIRs of two ears [2].
- (3) The ITD is estimated as a difference in the group delay of all-pass components, which is obtained by dividing an HRIR into a minimum phase component and an all-pass component [3].

However, all of these proposed methods have a problem. When a sound source is located on the contralateral side of the ear ("shadowed direction," hereafter), the accuracy of the ITD decreases. Therefore, in this article, a new estimation method is proposed to improve the accuracy of the ITD. In this method, HRIRs in adjacent directions are applied to calculate the cross correlation function.

2. Outline of proposed method

The detailed procedure of the proposed method is as follows:

- step: 1 Up-sampling all HRIRs (in this study, eight times [4]).
- step: 2 Search a lag point of the maximum of the cross correlation function between two HRIRs of adjacent azimuth angles. The lag point L is defined as

$$L = \{m | \max_{0 \leq m \leq N} r(m)\}, \quad (1)$$

where $r(m)$ is a cross correlation function with the lag point m . The cross correlation function $r(m)$ is calculated as

$$r(m) = \sum_{n=0}^N h(n+m, \theta_{a1})h(n, \theta_{a2}), \quad (2)$$

where $h(n, \theta_{a1})$ and $h(n, \theta_{a2})$ are the HRIRs of adjacent azimuth angles, and N is the length of each HRIR.

This process is applied to the HRIRs of each ear in all the elevation angles measured.

- step: 3 Search a lag point of the maximum of the cross correlation function between two HRIRs of adjacent elevation angles. The lag point is calculated in the same manner as step: 2; however, the cross correlation function is calculated for the azimuth angle of 0° only.
- step: 4 Revise the lags so that the lag average in all directions of one ear is equal to that of another ear. This is due to the assumption of the left-right symmetry of the head.
- step: 5 Translate all lags to delays from the reference HRIR, which is a typical HRIR for a front position in an azimuthal plane.
- step: 6 Calculate the ITD as the difference between the delays of each ear for each direction.

When the lag point at 0° , at the start of the procedure, is calculated up to step 2, the value is often unequal to that at the start of step 2 because of the accumulated error in the calculation of each lag point. In this study, the difference between the lags at 0° before and after step 2 was divided by the number of source azimuth angles, and then, this averaged value was added to each delay as an offset.

This proposed procedure has two advantages over conventional methods. First, a reliable correlation coefficient can be obtained because the correlation is always calculated using HRIRs in adjacent directions that are expected to have similar waveforms, even when the source is in a shadowed direction. Secondly, a fractional difference, i.e., a difference of less than one sampling time, between initial delays can be expressed by up-sampling HRIRs.

3. Experiment, results and discussion

3.1. Recording of HRIR

A set of HRIRs was recorded with a dummy head (KOKEN, SAMRAI) in an anechoic room. HRIRs were measured using loudspeakers on a spherical array. On this array, 72 loudspeakers (FE83E; Fostex Co.) were installed on

*e-mail: toyo@ais.riec.tohoku.ac.jp

†e-mail: kwatanabe@yamanashi.ac.jp

‡e-mail: iwaya@fir.riec.tohoku.ac.jp

§e-mail: yoh@ais.riec.tohoku.ac.jp

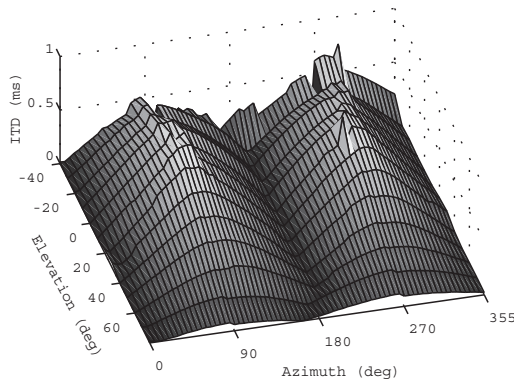
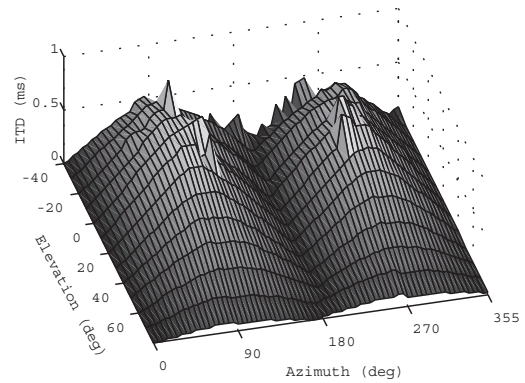
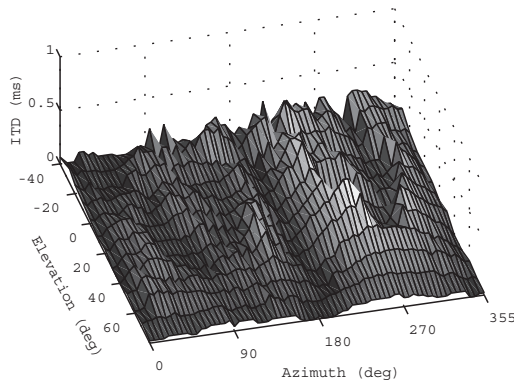
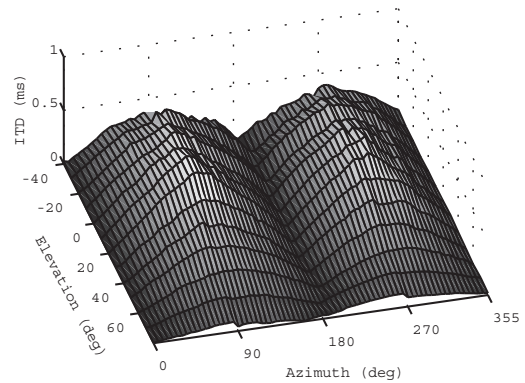
(a) ITD calculated by method proposed by Algazi *et al.* [1].(b) ITD calculated by method proposed by Kistler *et al.* [2].(c) ITD calculated by method proposed by Nishino *et al.* [5].(d) ITD calculated by method proposed by Minnaar *et al.* [3].

Fig. 1 ITD calculated by conventional method. The azimuth is 0° at the front position and increases clockwise. The elevation is 0° in the horizontal plane and increases upturn.

the ribs 1.5 m away from the center for every 10 degrees of elevation and azimuth angles. We can rotate the spherical array azimuthally with a precision of 0.1 degrees using a PC-controlled stepping motor. The intervals of the sound source position analyzed in the present study were 5° on the horizontal plane for all azimuth angles and 10° on the vertical plane for the angle range from -40° to $+80^\circ$. Thus, 936 (72×13) HRIRs were analyzed. The sampling frequency was 48 kHz, and the length of each HRIR was 512 points.

3.2. Results and discussion

The ITDs calculated by four different methods proposed in previous research studies [1–3,5] are shown in Fig. 1. Nishino *et al.* [5] argued only the initial HRIR delay, and the ITD estimation was not discussed. In Fig. 1, the initial delays of both ears are calculated by their method, and the ITDs are estimated as the differences between them. This figure shows that the results obtained by all the methods, except that developed by Nishino *et al.*, show relatively similar results. The results obtained by the methods proposed by Algazi [1] and Kistler and Wightman [2] show certain disturbances in shadowed directions, for example, when the azimuth is around 90° and 270° . In particular, when the sound source is on and near the horizontal plane, such disturbances are prominent. This could be attributed to the influence of the head diffraction that is strong near the horizontal plane. Minnaar's method shows relatively stable and reasonable results. In this sense, this method is superior to other three methods in

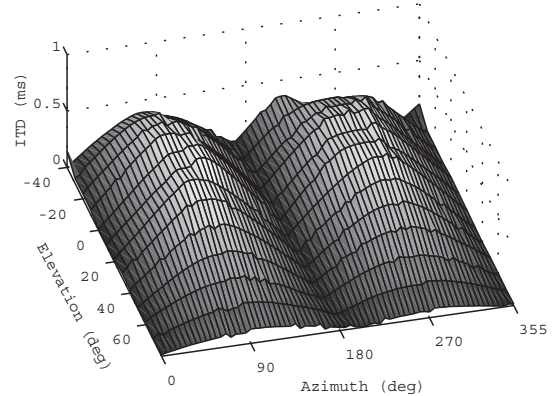


Fig. 2 ITD calculated by proposed method. The azimuth is 0° at the front position and increases clockwise. The elevation is 0° in the horizontal plane and increases upturn.

obtaining stable ITDs. However, one of the disadvantages of this method is its computational complexity in dividing an HRIR into minimum-phase and all-pass components.

Figure 2 shows the results obtained by the proposed method. The ITDs show a slightly larger disturbance at an elevation of -40° than those obtained by Minnaar's method (Fig. 1(d)). However, the overall results are very stable and similar to those obtained by Minnaar's method. In particular, on the horizontal plane, the overall disturbances obtained

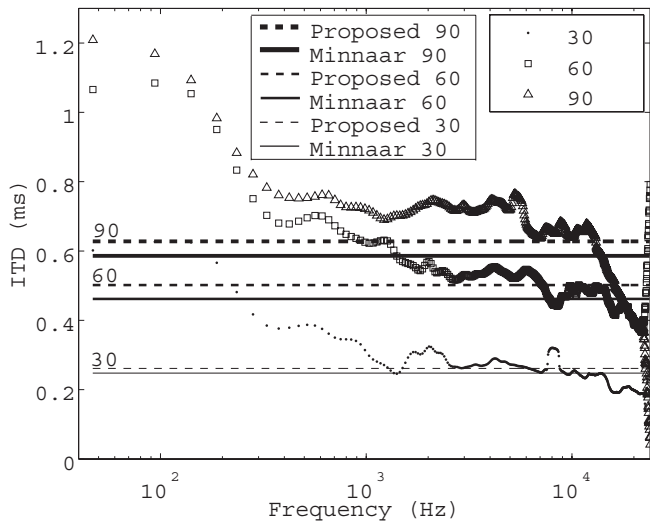


Fig. 3 ITD calculated from phase difference of HRIRs. This ITD is calculated at three azimuth angles in an elevation of 0° (●: 30° , □: 60° , △: 90°). Lines in this figure show Minnaar's ITD (solid line) and our ITD (dashed line). The widths of these lines indicate the azimuth differences, of 30° , 60° and 90° in thick order.

by our method at azimuths of around 90° and 270° are very small.

The ITDs obtained by our method and Minnaar's method are evaluated in the frequency domain. The plots in Fig. 3 show the ITDs calculated on the basis of the phase difference of HRIRs as a function of frequency. Moreover, the overall ITDs calculated by these two methods are shown by horizontal lines. The ITDs at the azimuth of 0° obtained by the two methods are similar. At the azimuths of 30° and 60° , the ITDs obtained by the proposed method are slightly larger than those obtained by Minnaar's method. Interestingly, the ITDs calculated by the proposed method fit well in the plots in the frequency ranges from 2 to 15 kHz for the azimuth of 30° and from 6 to 12 kHz for the azimuth of 60° . In contrast, the ITDs estimated by Minnaar's method do not fit well in the plots in any frequency range at the azimuths of 30° and 60° .

The ITDs calculated by our method fit the overall ITDs in the frequency range of 7–15 kHz, while those calculated by Minnaar's method fit in a frequency range higher than ours. This tendency is particularly distinct at the azimuth of 90° .

In this study, the intervals of source positions were only 5 degrees. The relationships between the intervals and for accuracy of the estimated ITDs were unclear. Watanabe *et al.* investigated an interpolation method for HRIRs and showed a high accuracy at the intervals of 10 degrees. In their method, the cross correlation function was used to estimate a lag between two HRIRs of adjacent azimuth angles. Given that the concept was the same, our method could be expected to show a high accuracy when the interval angle was 10 degrees.

4. Summary

In this article, we propose a new estimation method for ITDs. The ITDs estimated by this method showed an accuracy as high as that of the ITDs estimated by Minnaar's method, which is regarded as the best among previously proposed methods. The currently proposed method is advantageous because it is simpler than Minnaar's method in terms of the computational cost.

In a future work on our method, the influence of interval on the HRIR measurement must be investigated.

References

- [1] V. R. Algazi, C. Avendano and R. O. Duda, "Estimation of a spherical-head model from anthropometry," *J. Audio Eng. Soc.*, **49**, 472–478 (2001).
- [2] D. J. Kistler and F. L. Wightman, "A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction," *J. Acoust. Soc. Am.*, **91**, 1637–1647 (1992).
- [3] P. Minnaar, J. Plogsties, S. K. Olesen, F. Christensen and H. Møller, "The interaural time difference in binaural synthesis," *108th Conv. Audio Eng. Soc.*, preprint 5133 (2000).
- [4] K. Watanabe, S. Takane and Y. Suzuki, "Interpolation of head related transfer functions based on the common-acoustical-pole and residue model," *Acoust. Sci. & Tech.*, **24**, 335–337 (2003).
- [5] T. Nishino, S. Kajita, K. Takeda and F. Itakura, "Interpolation of the head related transfer function on the horizontal plane," *J. Acoust. Soc. Jpn.*, **55**, 91–99 (1999).