Geometrical evaluation of methods to approximate interaural time differences by broadband delays

Laurin Steidle[1], Robert Baumgartner[2]

[1] University of Vienna, 1010 Vienna, Austria, Email: laurin.steidle@univie.ac.at

[2] Acoustics Research Institute, Austrian Academy of Sciences, 1010 Vienna, Austria, Email: robert.baumgartner@oeaw.ac.at

Abstract

Interaural time differences (ITDs) constitute a prominent cue for the perceived lateralization of a sound source. Naturally occurring ITDs are slightly dispersive. However, psychoacoustic studies revealed that such natural ITDs may be indistinguishable from broadband delays that allow for more efficient implementations of virtual auditory displays. Consequently, numerous methods evaluating interaural cross-correlation, onset thresholding, or interaural group delay have been proposed to estimate the broadband delay best approximating natural ITDs. A recent study evaluated those estimators against listeners lateralization responses. Here, we extended this line of research by examining the geometrical consistency of proposed estimation methods by applying them to head-related transfer functions (HRTFs) of spherical heads for which the time of arrival can be determined analytically. With this procedure, we tested the estimators at various signal-to-noise ratios (SNRs) and used binaural stimuli to compare the estimates of standard approaches with predictions from a computational model including an approximation of the auditory periphery. Overall, approaches based on the maximum of interaural cross-correlation, especially when applied on signal envelopes, turned out to be generally most consistent with analytic geometric considerations and to be least affected by both additive and convolutional noise.

Introduction

The human auditory system utilizes a variety of acoustic cues to establish a spatial representation of a sound source. Head-related transfer functions (HRTFs) and the corresponding impulse responses (HRIRs) characterize the acoustic properties introduced by a listener's torso, head, and pinnae. The spatial separation between the two ears and the limited speed of sound cause interaural time differences (ITDs) whose magnitude systematically increases with spatial laterality with little dispersion. Concordantly, ITDs constitute a prominent cue for the perceived lateralization of a sound source, in particular if sounds provide energy at low frequencies. Psychoacoustic studies demonstrated that broadband interaural delays if chosen properly are perceptually equivalent to the naturally dispersive phase response [1, 2].

The simplicity of broadband ITDs offers a large potential to create very efficient auditory displays while raising the need for computational methods to estimate perceptually valid broadband ITDs from HRTFs. A large variety of such methods has already been proposed and recent work aimed to evaluate the correspondence of those ITD estimators against psychoacoustic responses [3]. They concluded that a simple -30 dB-thresholding procedure applied to the low-pass filtered impulse responses performed superior to all proposed methods including also a rather sophisticated model of auditory processing [4]. These investigations were only based on HRIRs with high signal-to-noise ratio (SNR).

In the present study we extended previous work by testing the geometrical consistency of the proposed methods at different SNRs and also if applied to binaural stimuli instead of impulse responses. To this end, we applied the ITD estimators to spherical head HRTFs, for which the time-of-arrival (TOA) at each ear and thus the ITDs can be derived analytically [5].

Methods

ITD estimators

In line with [3] we investigated ITD estimation methods following different approaches based on the analysis of onset thresholds, cross-correlations, or group delays – as listed in Table 1. All methods were evaluated both on broadband signals as well as versions low-pass filtered at 3 kHz by using a 6th-order Butterworth filter as in [3].

Table 1: ITD estimators

Method		Description
Threshold	(m)	Threshold of -10 dB relative to local peak
Cen-e2	(m)	Centroid of the squared envelope of HRIRs
MaxIACCr		Maximum of the IACC of HRIRs
MaxIACCe		Maximum of the IACC of energy envelopes of HRIRs
CenIACCr		Centroid of the IACC of HRIRs
CenIACCe		Centroid of the IACC of envelopes of HRIRs
CenIACC2e		Centroid of the squared IACC of energy envelopes of HRIRs
PhminXcor	(m)	Cross-correlation with minimum-phase version
IRGD		Integrated relative group delay
Dietz		Auditory lateralization model
	(m)	Monaural TOA estimators

HRIRs of spherical heads

We used numerically calculated HRIRs of spherical heads with various left-right-symmetric angular positions of the ears on the sphere, as provided by [5]. The head diameter was fixed to 87.5 mm and the 9 different ear positions are denoted in Figure 1. These HRIRs were sampled at 48 kHz and had a length of 5 ms. They were calculated for 1550 directions reaching from an elevation of -40° to 80° in steps of 5°. The azimuthal angle covered the full 360° with 2.5° steps within the interaural horizontal plane and

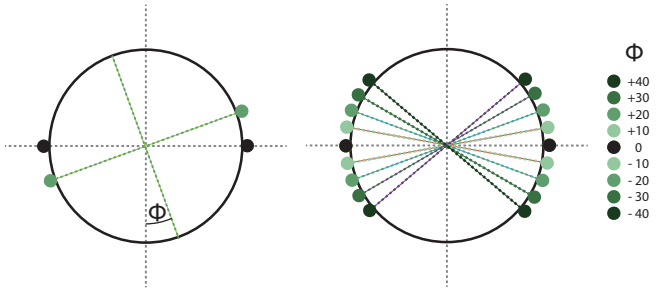

Figure 1: Angular positions of the ears on the spherical head. Φ denotes azimuth and Θ denotes elevation.

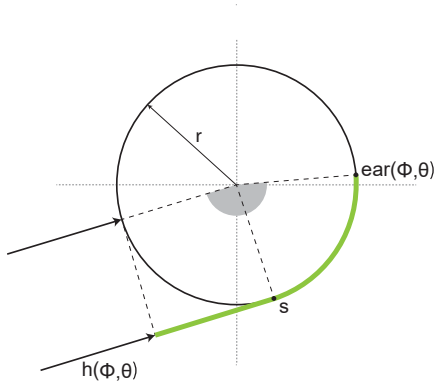

Figure 2: Illustration of the geometric model used to calculate TOAs. The circle represents the head with radius r. The green line shows the propagation path of the incoming wave $h(\Phi, \theta)$ to the ear position $ear(\Phi, \theta)$ used to calculate the TOA.

larger steps within other horizontal planes in order to obtain an approximately constant directional resolution.

Geometric model

The listener's head is modeled as a rigid sphere that is positioned at the center of the measurement system. Figure 2 illustrates the model for an arbitrary ear position on the sphere. Assuming plane-wave propagation, the TOA is modeled as the time the sound requires to travel along the shortest propagation path for a given direction h. Details on how this propagation path is calculated can be found in [5]. The ITD is simply the TOA difference.

Additive noise

For testing the robustness of ITD estimates against poor SNRs, we superimposed the reference HRIRs with Gaussian white noise. Tested SNRs ranged from 20 dB up to the clean signal (about 300 dB as limited only by numerical round offs).

Convoluted noise

While most of the ITD estimators were designed to be directly applied to HRIRs, the Dietz model includes a filterbank approximation of the auditory periphery and was designed to predict lateralization perception based on binaural stimuli. Hence, to allow fair comparisons between the standard estimators and this psychoacoustic model, tests were conducted on Gaussian white noise bursts with a duration of 0.5 s convolved with the respective HRIRs.

Several of the standard estimation models have fundamental issues when applied to stimuli. In particular, we excluded the PhminXcor method from this investigation because minimum-phase representations are only meaningful for linear time-invariant systems and not stimuli. Further, the threshold approach fails mainly because the reference peak may be located randomly across the whole stimulus duration. The Cen-e2 returns arbitrary results if the noise is much longer then the HRIR because the centroid of the convolution of the two then is dominated by the noise and therefore not anymore representative for the impulse response.

Performance measure

We evaluated the standard deviation between ITD estimates and the analytic ITDs, also referred to as the adjusted norm of residuals (ANR) [5]:

$$ANR = \sqrt{\frac{1}{N}\sum_i (\bar{y}_i - y_i)^2} \qquad (1)$$

Given a sampling rate of 48 kHz, the time difference between two adjacent sample points amounts to 20 μs. Therefore, an ANR of approximately 6 μs is considered as the lower bound reflecting best possible consistency.

Results

Effect of SNR

Figure 3a shows the ANR of the estimators for different SNRs. Note that the ANR is plotted on logarithmic scale. Overall, smaller SNRs yielded larger ANRs, that is, poorer performance. However, certain estimators are more severely affected by additive noise than others. Centroid-based methods performed rather poorly overall and were significantly degraded already at SNRs of 40 dB, while the MaxIACCe method performed best and almost identical for SNRs down to 30 dB. Also the other cross-correlation methods (MaxIACCr, PhminXcor) performed comparably well. The threshold and IRGD estimators performed well down to 40 dB but deteriorated at 30 dB or lower. At the lowest tested SNR of 20 dB, none of the methods performed well, resulting in ANRs $> 100\,\mu$s.

Low-pass filtering increased the perceptual validity of the vast majority of ITD estimators in a previous study [3]. For our geometrical examinations, low-pass filtering does not seem to be beneficial in general. While low-pass filtering generally improved consistency of Cen_e2 and CenIACCr, the opposite is true for both MaxIACC estimators. Also the other estimators yielded contradictory effects of low-pass filtering across SNRs.

Comparison with auditory model

Figure 3b shows the ANR for MaxIACCr, MaxIACCe, IRGD, and Dietz. For both broadband and low-pass filtered stimuli, MaxIACCe performed better than MaxIACCr, followed by the auditory model from Dietz et al. and finally the IRGD approach. The ANR of the IRGD approach fluctuated strongly, as reflected by very large standard deviations.

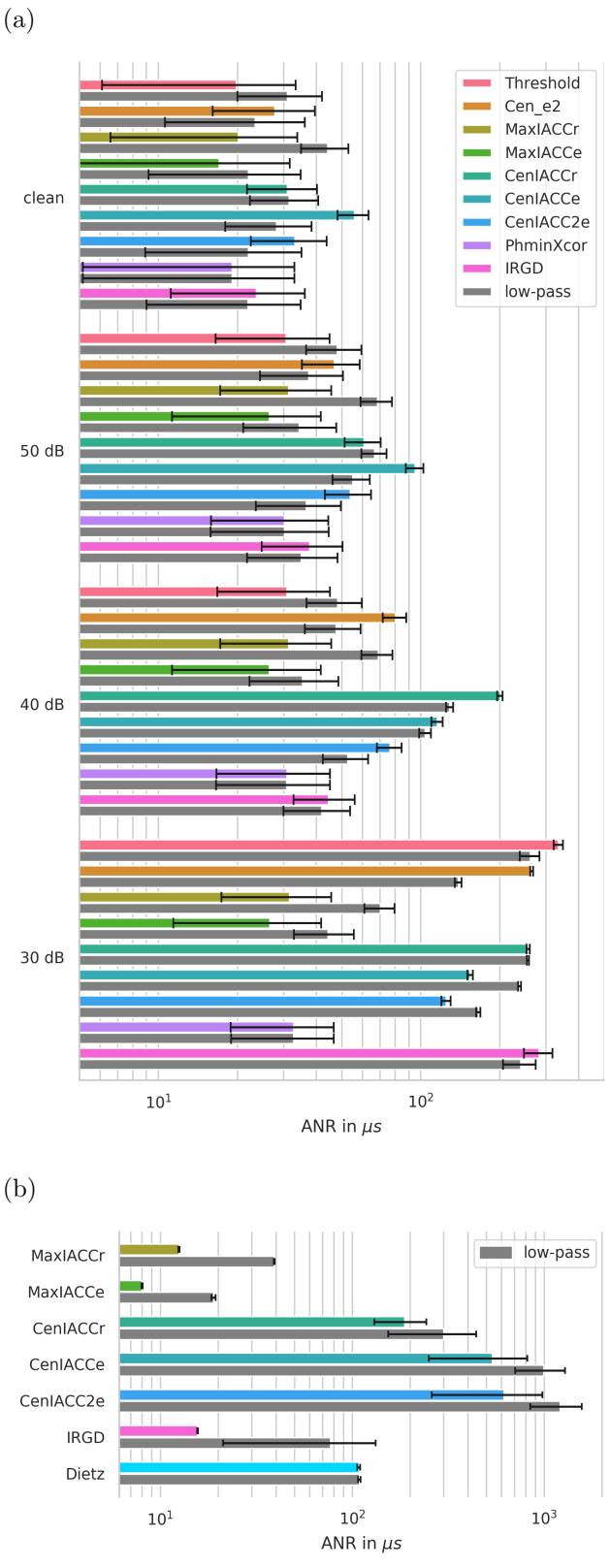

Figure 3: ANR of different ITD estimation methods (a) at different SNRs and (b) when applied to noise stimuli. The colored bars represent the ANR of the respective estimator based on broadband HRIRs. Error bars denote the standard errors of the means (across ear positions).

Discussion

Results of a recent psychoacoustic evaluation [3] ranked the following five estimators based low-pass filtered HRIRs as superior: Threshold $-30dB$ or $-20dB$, CenIACCr, MaxIACCe. In our evaluation, MaxIACCe and Threshold are the best estimators at high SNRs. Yet, CenIACCr performs poorly in our examination, especially for low SNRs, which might indicate a high SNR provided by the HRTF measurement procedure used in [3]. PhminXcor yielded very high geometrical consistencies in our evaluation, but performed only moderately well in [3]. The very simple approach of detecting onset thresholds at a predefined level of $-30dB$ relative to the global peak of the HRIR has been favored by [3] but has here been shown to be highly susceptible to noise for a relatively high threshold of $-10dB$. Lower thresholds yielded very inconsistent results for low SNRs and thus were not reported here.

Overall, our evaluations suggest that the MaxIACCe measure is geometrically most consistent and least affected by additive and convolutional noise. The code (`itdestimator` function) and data we used to conduct this study is publicly available as part of the open-source Auditory Modeling Toolbox (`http://amtoolbox.sourceforge.net/`) [6].

Acknowledgement

We want to thank the DEGA for the DAGA student grant, enabling us to participate at the DEGA.

References

[1] W M Hartmann and A Wittenberg. On the externalization of sound images. *J Acoust Soc Am*, 99(6):3678–88, 1996.

[2] A Kulkarni, S K Isabelle, and H S Colburn. Sensitvity of human subjects to head-related transfer-function phase spectra. *J Acoust Soc Am*, 105 (5):2821–2840, 1999.

[3] A Andreopoulou and B F G Katz. Identification of perceptually relevant methods of inter-aural time difference estimation. *J Acoust Soc Am*, 142(2):588–598, 2017.

[4] M Dietz, S D Ewert, and V Hohmann. Auditory model based direction estimation of concurrent speakers from binaural signals. *Speech Comm*, 53(5):592 – 605, 2011.

[5] H Ziegelwanger and P Majdak. Modeling the direction-continuous time-of-arrival in head-related transfer functions. *J Acoust Soc Am*, 135(3):1278–1293, 2014.

[6] P Soendergaard and P Majdak. The Auditory Modeling Toolbox. In Jens Blauert, editor, *The Technology of Binaural Listening*, pages 33–56. Springer, Berlin-Heidelberg-New York, 2013.