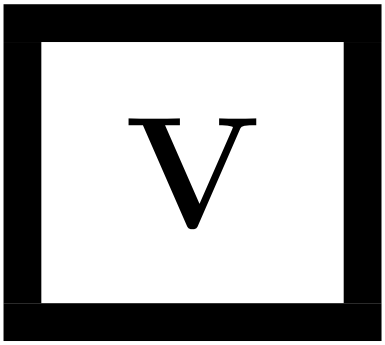

Publication V

Hirvonen, T. and Pulkki, V., "Interaural Coherence Estimation with Instantaneous ILD", in Proceedings of 7th Nordic Signal Processing Symposiun (NORSIG 2006), Reykjavik, Iceland, June 7-9, 2006.

INTERAURAL COHERENCE ESTIMATION WITH INSTANTANEOUS ILD

Toni Hirvonen and Ville Pulkki

Helsinki University of Technology
Laboratory of Acoustics and Audio Signal Processing
P.O. Box 3000 FI-02015 TKK Finland
Toni.Hirvonen@hut.fi

ABSTRACT

This paper presents a novel computational auditory model inspired by psychoacoustic and neurophysiological findings. The model utilizes the instantaneous difference between the two ear signals to predict spatial hearing cues that humans perceive. The main focus is on the interaural coherence cue, i.e. the perceived similarity between the waveforms of the ear signals. Simulations show that the proposed model is capable of predicting known psychoacoustical results.

1. INTRODUCTION

Computational auditory models are essential tools in psychoacoustic research. They simulate the auditory periphery, i.e. the outer-, middle-, and inner ear with great accuracy. In the upper stages however, where the two ear signals are combined in a so called binaural processor, the involved mechanisms are not completely understood. The majority of presently utilized binaural processor models are based on coincidence-counting mechanism introduced by Jeffress [1].

In a Jeffress-based scheme, the coincidence calculation is usually implemented as cross-correlation between the two ear signals. The cross-correlation function is commonly used to extract two important spatial hearing cues. The location of its maximum peak in the delay axis can be interpreted as the interaural time difference (ITD) of the incoming sound between ears. Furthermore, the maximum height of the normalized function is used to describe interaural coherence, or similarity between the ear signals. However, it has been suggested that the human nervous system could not perform the required cross-correlation normalization with sufficient accuracy [2]. Also, the physiological validity of the Jeffress model has been recently questioned [3].

A number of studies have shown that the ITD mechanism is quite sluggish, with a time constant as high as 250 ms, and that the interaural level difference (ILD) decoding is much faster [4] [5]. While the exact time resolution of the ILD mechanism has not been measured, we assume in this paper that the small integration time of the ILD makes it possible for humans to perceive instantaneous ILD, or the difference between ear signals, as opposed to the common viewpoint where ILD is merely thought as an overall level difference averaged over a relatively long time period. A novel auditory model that is based on calculating the instantaneous ILD is proposed. We examine how well the model can be used in predicting interaural coherence detection.

This work was supported by The Academy of Finland (projects no. 201050 and 105780)

2. ILD MODEL FOR LOCALIZATION AND BINAURAL COHERENCE DETECTION

2.1. Fundamentals of ILD Processing

A simple hearing scenario where a single point-like source radiates sound from far-field (>2 meters) illustrates the basic ILD phenomenon: if the source is located on the side of the listener, the shadowing of the head causes attenuation in the sound arriving at the opposite ear. The attenuation is minor at low but notable at higher frequencies. If the source is located close to the head (near-field), ILD increases. Despite its frequency-dependent nature, humans can decode ILD at all frequencies within 1 dB. For example, large ILDs affect localization performance at low frequencies [6]. This is counterintuitive in the sense that large low-frequency ILDs do not occur with most natural sound sources.

2.2. Decoding Coherence with ILD

Everyday listening in a reverberant environment often includes situations where the ear signals differ greatly from each other. Coherence is in spatial hearing research defined as the similarity between the ear signals, and is often calculated as the maximum of the interaural cross-correlation. It has been determined as a perceptually important factor in e.g. feeling of spaciousness [7]. In the case of non-coherent ear signals, the resulting instantaneous ILD understandably fluctuates rapidly. Because of the speed of the ILD detec-

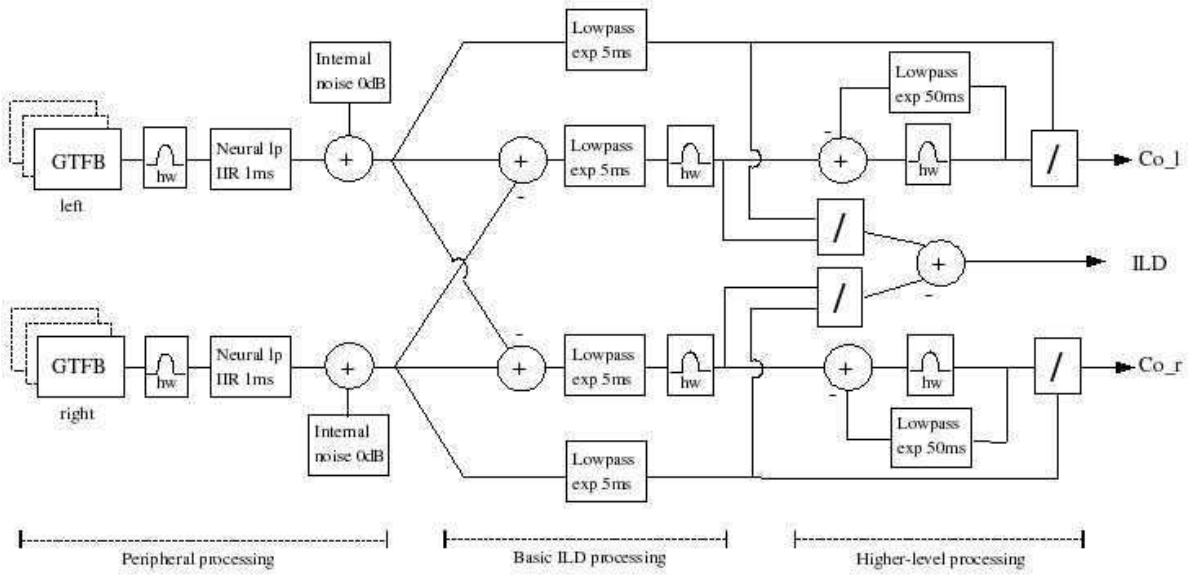

Fig. 1. *The proposed model structure for estimating the ILD localization and coherence cues. (See Section 2.3 for details.)*

tion, we hypothesize that it is possible that coherence is at least partly detected using instantaneous ILD information.

2.3. Structure of the Model

This section presents a computational implementation, where instantaneous ILD is used to estimate coherence and ILD localization cue. Fig. 1 shows a flow-chart of the model. The design is based on the known physiological structure of the human hearing system. However, the approach presented here is mainly functional.

The first stages of the model, starting from the left-hand side, mimic the well-documented peripheral functions of the cochlear nucleus: gammatone filterbank (GTFB) divides the signal into critical frequency bands that are processed separately in accordance with the human frequency resolution. A Matlab implementation by Slaney was used for this purpose [8]. The transformation of vibrations into neural impulses is here modeled with a simple half-wave rectification (hw), which is also employed in the further stages as neural signal cannot be negative. The neural lowpass IIR filter (τ=1 ms) models the asynchrony of the neural transformation: the peripheral neural impulses remain synchronous to the input waveform only up to approximately 1000-2000 Hz. All filters utilized in the model are normalized to have a maximum gain of 0 dB. Gaussian white noise is added to the signal at 0 dB level to simulate the internal noise of the neural system.

The parts following the internal noise addition represent the novel approach of this paper and model functionally the basic ILD processing that takes place in the Lateral Superior Olive (LSO), as well as other organs. LSO is one of the first sites of binaural interaction located in the Superior Olivary Complex (SOC) and it receives inputs from both cochlear nuclei [9]. LSO is commonly regarded as the initial and most important stage of encoding of ILDs with most of its cells being inhibition-excitation (IE) type [3]. IE here refers to a process where the contralateral ear signal inhibits the ipsilateral input.

It should be noted that although LSO cells have characteristic frequencies mostly above 1000 Hz, no high-pass filtering etc. is used here for the reason that humans can also detect ILD at low-frequencies. We approximate the LSO cells with a simple subtraction: both ear channels are subtracted from each other sample-by-sample to calculate the instantaneous ILD. The following lowpass filter (convolution with an exponentially decaying series, τ=5 ms) functions as a temporal integrator and simulates the slowness and saturation inside LSO cells. After another half-wave rectification, the signal information is used to extract psychoacoustically relevant information similarly as in the higher stages of auditory processing. The present implementation includes two specific "paths": ILD-, and Co-channel.

The Co-channels in Fig. 1 are used to estimate the perceived coherence cue. The coherence channel outputs were designed to increase as the correlation between the ear signals decreases and humans are more probable to perceive it. The processing includes a negative feedback loop with a lowpass filter (τ=50 ms). This mechanism is used to remove the steady DC-component of the instantaneous ILD. In an anechoic environment, a sound coming from non-zero azimuth causes a constant, non-varying ILD, which in the

model manifests itself as DC-component in the Co-channel. Thus it is appropriate to remove this, focusing only on the time-varying component.

Both the ILD and Co-channels make use of the neural signal prior to the basic ILD-subtraction stage: Both ear signals are divided sample-by sample with the "mono" signal filtered with the same filter as used in the temporal integrator. This is done for sake of normalizing the Co- and the ILD-outputs as relative to the input and between frequency channels.

The ILD-output information is intended to be used directly as the ILD localization cue. After the normalization division, both ear ILD-signals are combined to a single ILD by subtraction. This yields the instantaneous ILD cue in the middle output of Fig. 1.

3. SIMULATIONS

This section presents tentative simulations that compare the model outputs to existing psychoacoustical data. It should be noted that the parameters of the model were not meticulously adjusted and can probably be tuned to yield better results.

3.1. Coherence as a Function of Interaural Correlation and Frequency

The first simulation investigates if the model can estimate interaural coherence similarly as humans do. The Co-output is monitored when the model input signal is noise at 70 dB level, with various degrees of interaural decorrelation. Culling *et. al.* have performed a similar test with human test subjects [10]. The purpose here is to see how the model predictions correspond with the psychoacoustical results. The mean output values of both ear signals were used to calculate the total coherence cue as: $mean(Co_l) * mean(Co_r)$. As in [10], signal-to-noise ratio (SNR) is calculated as a function of interaural correlation ρ as:

$$SNR = 10 * log_{10}((1-\rho)/(1+\rho)), \qquad (1)$$

so that 0 dB indicates fully decorrelated ear inputs.

Fig. 2 shows the results of the simulation. The values are normalized so that the maximum of the results is set to 0 dB. It can be seen that the mean correlation cue maximum value occurs at small interaural correlation values and at low frequencies. This coincides with the results presented in Figs. 6 and 7 of [10]. The similarity is limited to general trends, as the comparison between the model output and the detection index used in [10] is not directly possible.

The frequency and SNR value limits in Fig. 2 were chosen according to those used in [10]. The CO-output was also examined at higher frequencies, and it was found that the simulations show similar phenomena as the results for

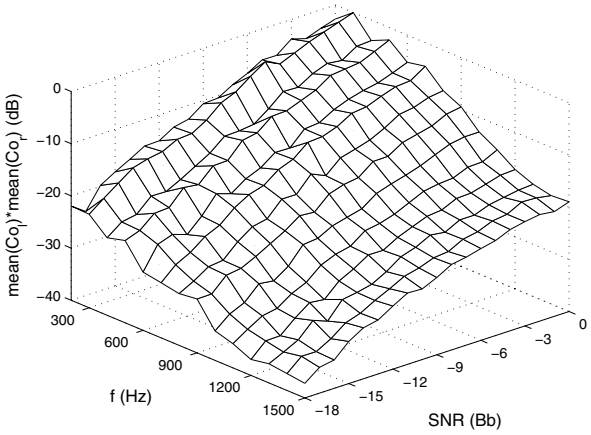

Fig. 2. *Normalized mean coherence output cue as a function of SNR (i.e. correlation) and frequency with interaurally correlated noise.*

normalized correlation detection presented in Fig. 1 of [11]. A detailed presentation of these results is left to the future.

3.2. ILD and Coherence as a Function of Azimuth and Frequency

The purpose of this simulation is to investigate how the the model responds to different azimuth incidence angles of the input sound. A white Gaussian noise sample with 70 dB level was filtered with measured dummy-head head-related transfer functions (HRTFs) [12]. Due to their symmetry, only the measurements from the right-hand side (azimuths 0-80°) were used. Each critical frequency band is processed separately. Thus, the model estimates the coherence and the ILD localization cues as a function of azimuth angle and frequency.

To obtain the ILD cue, the ILD output is simply averaged over the signal duration at each azimuth-frequency value. Figure 3 shows the obtained values. The output ILD value is normalized so that 1 indicates the ILD cue pointing directly to the other side of the head. It can be seen that the mean ILD output behaves similarly as the level difference between ear canal signals, increasing with frequency and incidence angle and not being large at low frequencies. At approximately 1 kHz there is a notable notch, similarly as with real measurements.

Fig. 4 shows the mean of both Co-channels calculated as in the previous section. Values were normalized to the maximum value in Fig. 2. It can be seen that the resulting values are relatively small, rising slightly with the increasing incidence angle. Generally, the maximum here is 20 dB smaller than in the experiment using non-coherent noise. This result is desirable in this simulation, because as point-like sound

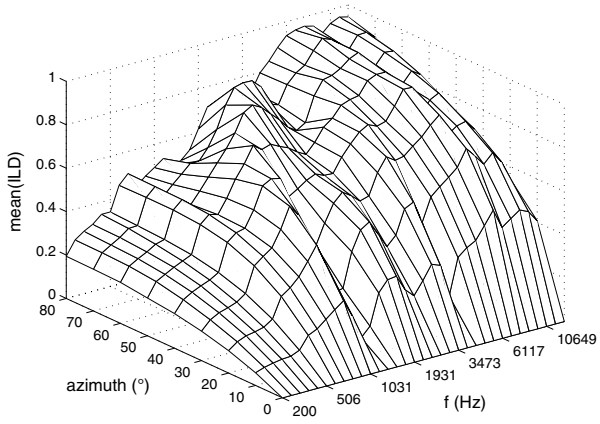

Fig. 3. *Mean ILD output cue as a function of simulated azimuth angle and frequency with HRTF-filtered noise.*

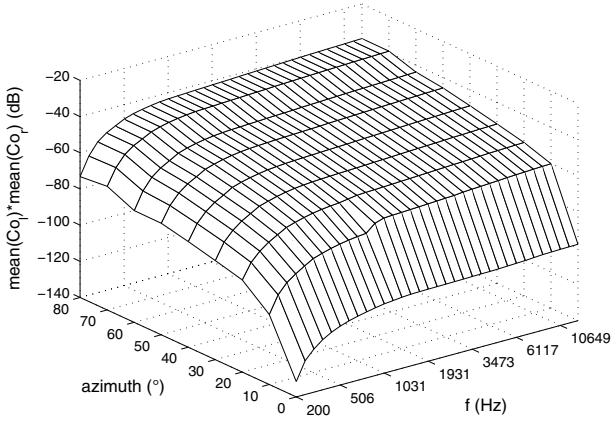

Fig. 4. *Mean coherence output cue as a function of simulated azimuth angle and frequency with HRTF-filtered noise. Values have been normalized to the maximum value in Fig. 2.*

sources were simulated, the resulting coherence cue should be low.

4. DISCUSSION

The equalization-cancellation (EC) model proposed by Durlach [13] bears some resemblance to the present implementation in the sense that it is also based on the subtraction of the two ear signals. However, the approach is somewhat different as the EC-model included specific delay and level compensation before signal subtraction and is usually applied to binaural detection tasks. Also, Reed and Blum have proposed an LSO-based ILD model [14], but it does not consider the perceived coherence.

5. REFERENCES

[1] L. A. Jeffress, "A place theory of sound localization," *J. Comp. Physiol. Psych.*, vol. 41, pp. 35–39, 1948.

[2] S. van de Par, C. Trahiotis, and L. R. Bernstein, "A consideration of the normalization that is typically included in correlation-based models of binaural detection," *J. Acoust. Soc. Am.*, vol. 109, no. 2, pp. 830–833, 2001.

[3] B. Grothe, "Sensory systems: New roles for synaptic inhibition in sound localization," *Nature Neuroscience*, vol. 4, pp. 540–550, 2003.

[4] D. W. Grantham, "Discrimination of dynamic interaural intensity differences," *J. Acoust. Soc. Am.*, vol. 76, no. 1, pp. 71–76, 1984.

[5] T. N. Buell and E. R. Hafter, "Discrimination of interaural differences of time in the envelopes of high-frequency signals: Integration times," *J. Acoust. Soc. Am.*, vol. 84, pp. 2063–2066, 1988.

[6] D. S. Brungart, N.I. Durlach, and W. M. Rabinowitz, "Auditory localization of nearby sources ii: Localization of a broadband source," *J. Acoust. Soc. Am.*, vol. 106, no. 4, pp. 1956–1968, 1999.

[7] M. Barron and A. H. Marshall, "Spatial impression due to early lateral reflections in concert halls: the derivation of a physical measure," *J. Sound Vib.*, vol. 77, no. 2, pp. 211–232, 1981.

[8] M. Slaney, "An efficient implementation of the Patterson-Holdsworth filter bank," Tech. Rep. 35, Apple Computer, Inc., 1993.

[9] W. A. Stotler, "An experimental study of the cells and connections of the superior olivary complex of a cat," *J. Comp. Neurol.*, vol. 100, pp. 401–423, 1953.

[10] J. F. Culling, H. S. Colburn, and M. Spurchise, "Interaural correlation sensitivy," *J. Audio Eng. Soc.*, vol. 110, no. 2, pp. 1020–1029, 2001.

[11] L. R. Bernstein and C. Trahiotis, "The normalized correlation: Account for binaural detection across center frequency," *J. Acoust. Soc. Am*, vol. 100, no. 6, pp. 3774–3784, 1996.

[12] B. Gardner and K. Martin, "HRTF measurements of a KEMAR dummy-head microphone," Tech. Rep. #280, MIT Media Lab Perceptual Computing, MA, USA, 1994.

[13] N.I. Durlach, "Equalization and cancellation theory of binaural masking-level differences," *J. Acoust. Soc. Am.*, vol. 35, no. 8, pp. 1206–1218, 1963.

[14] M. C. Reed and J. J. Blum, "A model for the computation and encoding of azimuthal information by the lateral superior olive," *J. Acoust. Soc. Am.*, vol. 88, no. 3, pp. 1442–1453, 1990.