VII

Publication VII

Pulkki, V. and Hirvonen, T., "Computational Count-Comparison Models for ITD and ILD decoding", ICA 19th Int. Congress on Acoustics, 2007.

19th INTERNATIONAL CONGRESS ON ACOUSTICS
MADRID, 2-7 SEPTEMBER 2007

COMPUTATIONAL COUNT-COMPARISON MODELS FOR ITD AND ILD DECODING

PACS: 43.64.Bt

Pulkki, Ville; Hirvonen, Toni;
Lab. Acoustics and Audio Signal Processing, Helsinki University of Tech., POBox 3000, FI-02015, Finland

ABSTRACT
Recent neurophysical studies suggest that binaural decoding is based on count comparison in cases of both ITD and ILD decoding. In such mechanisms, the neural signals are stronger in the auditory pathways leading to the ipsilateral hemisphere when a signal is presented earlier, or with higher level, to the contralateral ear. This paper describes a simple computational model implementing binaural cue decoding based on count-comparison principles. In the model, ITD and ILD are decoded using separate mechanisms, inspired by the functions of Medial Superior Olive and the Lateral Superior Olive found in the mammal brainstem. It is also assumed, based on psychoacoustic data, that ITD decoding is sluggish, and ILD decoding is faster. It is shown that the proposed mechanisms decode static ITD and ILD similarly as humans do. In addition to ITD and ILD, humans can also detect interaural coherence, or the similarity between ear signals. Interaural coherence is in the model decoded as the range of temporal variations of ILD.

INTRODUCTION
The human, like many other mammals, is capable of perceiving the direction and the distance of a sound source. These attributes are decoded using mainly interaural differences [1]. Humans are also able to detect the similarity of sound signals in the ears, which leads to perceiving the extent of a sound source, and the attributes of room reverberation. Both the neurophysiology and the psychoacoustics of binaural hearing have been studied extensively during the last decades, and many psychophysical binaural phenomena have been explained using computational models. The models have also been bolstered with neurophysiological data. One of the most influential models has been the Jeffress model of ITD decoding [2], which assumes that the neural impulses from the ears propagate to coincidence counting neurons, which in turn fire if an impulse arrives from both ears at the same time. The propagation delay to neurons from each ear varies systematically, which enables the azimuthal position of sound source to be decoded based on the place of most active neuron on the neuron array. Interaural coherence (IAC, similarity between the ear signals) is in the Jefferess model detected based on the relative output rate of the most active neuron. There are also some neurophysiological studies in which such activity has been found in the brains [3].

However, in recent studies some neural mechanisms which do not match with the Jeffress model have been found. As reviewed by Grothe, these results seem to fit better to a count-comparison model for localization [4], where the output of the MSO organ in ipsilateral hemisphere is larger when the sound arrives earlier to the contralateral ear. The output of the organ would then be compared to the activity of the contralateral organ. There exists publications both supporting and questioning the suggestions made in [4]. In this work we are investigating if a count-comparison model of binaural decoding can be used explain some psychoacoustic test results. If simulations with such a model would fit the psychoacoustical data, it could be taken as a evidence supporting the proposed decoding principle.

AUDITORY PATHWAY
In this section, the knowledge of neurophysiology and neuroanatomy used in this work is reviewed shortly. The sound arriving to the inner ear is transferred into neural impulses in the cochlea, which is tonotopically organized. The neural paths departing from cochlea are thus frequency selective. The neurons show the largest activity at their corresponding best frequencies. From the cochlea, signal traverses into cochlear nucleus (CN), from where it is routed to different organs. The neurons which project to organs decoding binaural cues have two types. Some of them synchronize to the phase of signal very precisely (phase-locking neurons) at frequencies below approximately 1-2 kHz, and some of them synchronize precisely to onsets and offsets of sound events (transient-locking neurons) [5].

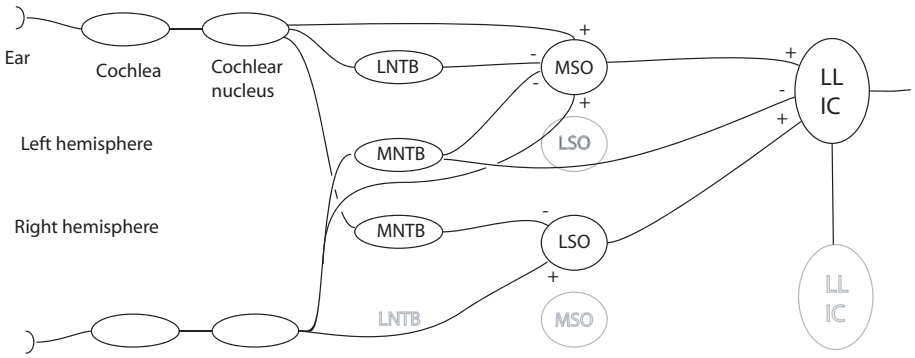


Figure 1: The most important neural connections found in the mammalian hearing system for binaural decoding below midbrain. For simplicity, only the connections are shown which are leading to the IC in the left hemisphere.

The projections important in binaural cue decoding lead from CN to organs located in the brainstems of both hemispheres, which are shown in Fig. 1. Excitatory terminations can be found in the Medial Superior Olives (MSO) of both hemispheres, and in the ipsilateral Lateral Superior Olive (LSO). CN also projects to the contralateral medial nucleus of trapezoid body (MNTB) and to the ipsilateral Lateral Nucleus of Trapezoid Body (LNTB), both of which further provide phase locked inhibition to the organs they are connected to [6, 4]. The LNTB projects solely to the MSO, and the MNTB has three main ipsilateral projections, the MSO, the LSO and the Ventral Nucleus of Lateral Lemniscus (VNLL) [7]. The VNLL is situated near Inferior Colliculus (IC) in the midbrain, thus, for simplicity, the VNLL connection is drawn to organ labeled as LL / IC

The MSO thus receives both excitation and inhibition from both hemispheres, and it is assumed to decode primarily ITD [4]. The outputs of guinea pig MSO neurons have been recorded using white noise with different ITDs [8]. The results indicate that the earlier the contralateral signal arrives with respect to the ipsilateral signal, the higher the output rate of the MSO. The highest rate is obtained, when the delay between signals is 1/8 of the period of the best frequency of the neuron. Grothe assumes [4] that the MSO neurons act as coincidence counters, which receive both excitation and inhibition originating from both hemispheres. The functioning of the MSO can then be explained by assuming that the contralateral inhibition arrives before excitation, which effectively delays the contralateral excitation, and makes the action potential grow slower. Grothe also suggests that ipsilateral excitation arrives before inhibition, which effectively shortens the excitation. With this mechanism, the delaying of a sound to ipsilateral ear would make the output higher up to certain value of the delay.

The LSO receives excitation from ipsilateral CN, and inhibition from contralateral CN, via MNTB [9]. It is assumed to decode ILD in a way that the stronger ipsilateral signal is compared to contralateral signal, the stronger is the output of LSO. LSO projects most prominently to the contralateral IC, in contrast to the ipsilateral IC projection of MSO. However, this can be understood due to the fact that when sound source is in the contralateral hemisphere, the increased activity of both MSO and LSO will be projected to the same IC. LSO has also found to respond very fast to ILD between ear signals [9].

There also exist some psychoacoustical data on sluggishness of binaural cue decoding which is of interest here. Some studies suggest that ITD processing is relatively sluggish, and ILD processing is decoded prominently faster [10]. The fast ILD decoding would enable the decoding of IAC using ILD cue, since when the ear canal signals are not equal, and when the ILD is decoded instantly, there will naturally be strong ILD fluctuations. This is in accordance with [11], where it is suggested that interaural phase and level fluctuations are used in binaural detection instead of binaural cross correlation.

MODEL

The auditory model constructed in this work attempts to follow the neurophysiological knowledge. However, the model tries not to imitate the functioning of the organs precisely. The target of this work is to test if the modeling concept can be used to explain some basic psychoacoustical results. The suggested model is presented in Fig. 2.

Cochlea and cochlear nucleus

The cochlea is modeled simply by using a gammatone filterbank (GTFB) [12]. The phase-locking and neu-

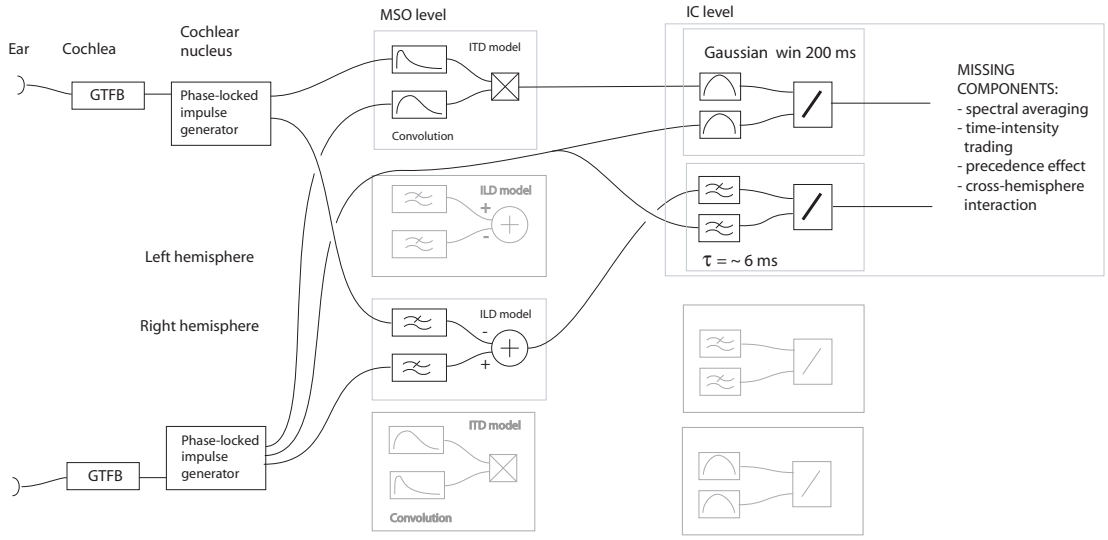


Figure 2: Computational model of binaural decoding. For simplicity, only the connections leading to the left hemisphere are shown.

rons of CN are modeled with a simple heuristic model. The signal is first half-wave rectified, after which the halfwaves are replaced with impulses preserving the energy of signal. The impulses are inserted to the time instants where the rectified signal crosses zero after being positive. The impulses are, in the used sampling rate of 48000 Hz, unrealistic in the sense that they are too short to actually occur in the brains. However, in the MSO and LSO models, the impulses are convolved with longer responses or low-pass filtered. This process can be interpreted to also implement the finite temporal response of the CN neurons, as well as the lose of synchronization to the input sound, which occurs at frequencies above 1-2 kHz. The transient-locking neurons of the CN are also modeled with a heuristic method; if the previous and next impulses in time have smaller amplitude than the present impulse, the amplitude of present impulse is magnified by a factor a. The factor has been set typically to the value of 10 in this work.

MSO modeling

In the present model, MSO acts as a simple multiplier, which receives input from both hemispheres. The input from CNs is convolved with a response, which is different for the inputs from each hemisphere. The contralateral input is convolved with an exponential rise and decay, having temporally relatively long slopes. The input from the ipsilateral side is convolved with a function having faster exponential rise and decay. These different ipsilateral and contralateral responses can be interpreted to implement Grothe's suggestion of the arrival orders of excitation and inhibition from both sides to MSO [4]. Since the excitation is assumed to arrive after inhibition in the case of contralateral input, the slopes of rise and decay are temporally longer. Correspondingly, the excitation arrives before inhibition with the ipsilateral input, which is here modeled as a single, temporally very short response.

The parameters of the functions were tuned by comparing the output of the MSO model to recordings of guinea-pig MSOs conducted in [8]. In the test, white Gaussian noise was presented with static ITDs of different values. The same situation was simulated with the model, and the result can be seen in Fig. 3. The parameters were hand-tuned, and a good correspondence between the neurophysiological data and the model output was found when the ipsilateral pulse was very fast, having an exponential onset of 4μs, and a fast decay with a time constant of 6 μs. The time constant for the contralateral response rise was adjusted to 10%, and response length to 26% of the period of center frequency of corresponding frequency band. Correspondingly, the time constant for the contralateral response decay was set to 10% of a period.

It can be seen that the model output matches well with the results of from the neurophysiological recordings near the ITD value of zero. The MSO output is plotted with two values of a, 1 and 10. With value of 1, the neurons of the CN sensitive to transients are not responsive at all, and with value of 10, they produce a prominent response. It can be seen that the value of a has an effect only on sidelobes of the response. With the value of 1, the simulated response departs largely from the recorded response, and with the value of 10, the simulated response is relatively close to the neurophysiological response.

LSO modeling

LSO has been found to include bipolar cells receiving excitation from the ipsilateral CN, and inhibition from

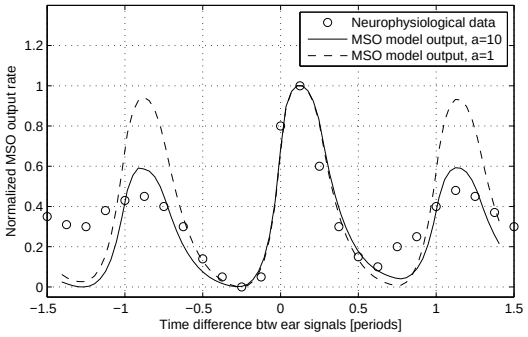


Figure 3: The output of the MSO model with frequency channel at 775 Hz compared to neurophysiological data adapted from [8]. The parameter value $a = 1$ refers to a situation where the transient-sensitive neurons are not effective, and $a = 10$ refers to a situation where they are effective.

the contralateral CN. Thus, LSO is modeled as a simple subtractor, as shown in Fig. 2. Prior to subtraction, the responses arriving from the CN are low-pass filtered with a fourth-order IIR filter having a cut-off frequency at 400 Hz, which effectively implements the lose of neural synchronization after 1-2 kHz. Although some neural recordings of LSO outputs are available, the output was not tuned according to them in this work.

Lateral lemniscus and Inferior Colliculus

The neurophysiological functions of these organs is less well known. However, all impulses originating from CN travel through them. The activity related to spatial hearing is of interest here, and we hypothesize certain processes happening in LL and IC.

Both MSO and LSO model outputs depend on input signal level. Higher input causes higher output, and vice versa. To reflect the position of sound sources, the output levels have to be normalized somehow. In traditional implementation of count-comparison models, the output of MSO and LSO models are compared between hemispheres. However, in lesion studies it has been found, that unilateral lesions above MSO level cause deficits only for localization in contralateral side [13], which implies that the comparison to contralateral side is not essential part of processing. It has been found that ipsilateral MNTB provides inhibitory input to VNLL [7]. This makes us to assume that the outputs of MSO and LSO are normalized using one of the inputs of MSO and LSO, respectively, more precisely, with the MNTB output. This would explain how the localization ability can be lost only in one hemisphere.

In practise, the outputs of the MSO and LSO models arriving to LL and IC are divided by the contralateral signal after temporal averaging, as shown in Fig. 2. We also assume that some bilateral comparison occurs in midbrain level. However, this process has not been implemented yet. In this model, temporal averaging is implemented by simply convolving the MSO output with a Gaussian time window with a length of 200 ms. The LSO output is simply slowed down by filtering it with a first-order IIR having temporal constant of 6 ms.

We also assume that the mechanisms producing the precedence effect [14] are situated mostly in LL / IC. The precedence effect is a mechanism affecting a sound source to be localized based only on the direction of direct sound, and not on directions of early reflections. The proposed bilateral connections between ICs and LLs as shown in Fig. 1, could explain mechanism. However, this part of the model has not been implemented yet.

THE RESPONSE OF MODEL TO SIMPLE BINAURAL LISTENING CONDITIONS

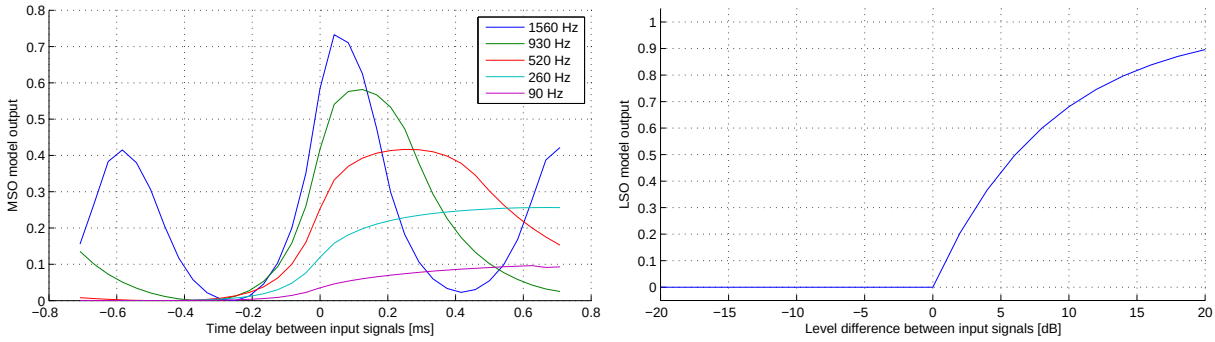
In this section, we will present the responses of the model to some simple binaural scenarios.

Constant time difference

In this simulation, white noise bursts with length of 100 ms were presented to the model with different ITDs. As could be expected, only the MSO model output is significantly dependent on ITD, and the output of the LSO model is low at all frequencies and with all ITDs. Thus, only the response of the MSO model is shown in Fig. 4 a. At low frequencies, the output depends monotonically on ITD, and at higher frequencies the output has cyclic nature. This fits at least qualitatively to psychoacoustic results, since the lateralization of a sinusoids depends circularly on the phase difference between the ears [1].

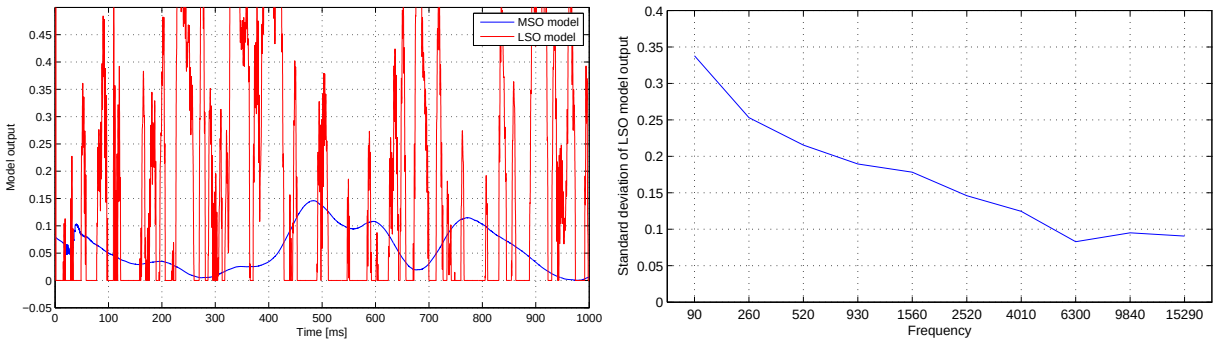
Constant level difference

In this simulation, the stimulus was a coherent Gaussian white noise burst of 100 ms. The ILD of the noise



a) b)

Figure 4: a) The response of the MSO model to coherent Gaussian white noise input with constant time delay between ears at different frequency bands. b) The response of the LSO model to coherent Gaussian white noise input with constant level difference between ears.



a) b)

Figure 5: a) The temporal response of the LSO and MSO models to incoherent white noise at frequency band with center frequency 520 Hz. b) The standard deviation over time of the output of the LSO model to incoherent white noise input as function of frequency.

was varied between -20 dB and 20 dB. The results show that the MSO model output does not change with ILD, which was expected. As also anticipated, the LSO model output varies monotonically with ILD. The output was the same at all frequencies, which is shown in Fig. 4 b. This corresponds to psychoacoustic studies, according to which ILD is decoded with the same accuracy at all frequencies [1]. In our previous work, an earlier version of the LSO model has also been tested with HRTF-simulated sources [15], where it was found, that the LSO output carried the information of ILD between ear signals generated by real sources.

Incoherent input signals

As mentioned, humans can also decode the similarity between ear canal signals, which is in this paper denoted loosely as IAC. Humans are most sensitive to IAC at low frequencies, and the sensitivity decreases with frequency. Traditionally, it has been assumed that humans would decoded IAC with a Jeffress-type mechanism. As Jeffress-type mechanisms are in this study assumed not to be physiologically prevalent, and count-comparison models have been suggested instead, a question arises how would IAC be decoded using count-comparison models. This was tested simply by simulating a listening condition where independent white noises were used as inputs to different ears, in which case the ear canal signals are different. The resulting cues for different frequency channels were monitored as function of time. It was found that the MSO output varied relatively slowly, whereas the LSO model produced rapid temporal changes. This is mainly due to the differences in effective window lengths in the temporal integration. The LSO and MSO outputs at the frequency channel with a center frequency of 520 Hz are plotted in Fig. 5 a, where this behavior can be seen. When this result was monitored at different frequency bands, it was noted that the variance of the LSO output depends on frequency. This variance is quantified in Fig. 5 b by plotting the standard deviation of the LSO output over time as function of frequency. It can be seen that the deviation decreases with frequency, which agrees with psychoacoustic results [1].

This shows that the output of the presented model carries information about IAC in temporal variations of LSO, as also suggested in [11]. This was shown already in [15] with earlier version of the LSO model.

Additionally, it has been shown that the model can be used to explain basic binaural masking level difference results, and some binaural pitch results [16].

DISCUSSION

An implementation of a count-comparison model for binaural cue decoding was presented in this paper. The parameters of the model were tuned partly using neurophysiological data and partly by using psychoacoustical data. The model was tested with a few simple binaural listening simulations. The results from these simulations fit psychoacoustical data at least qualitatively, and the count-comparison modeling approach seems valid in modeling of psychoacustical listening tests. However, further work is needed to develop the model. For example, the model currently lacks the interaction between hemispheres that is assumed to occur at the level of the midbrain, which could explain some psychoacoustical issues, such as the precedence effect and binaural time-intensity trading.

ACKNOWLEDGMENTS

Ville Pulkki has received funding from the Academy of Finland (project 105780) and from the Emil Aaltonen foundation.

References

[1] J. Blauert. *Spatial Hearing*. The MIT Press, Cambridge, MA, USA, revised edition, 1997.

[2] Lloyd A. Jeffress. A place theory of sound localization. *J Comp. Physiol. Psychol.*, 41:35–39, 1948.

[3] P. Joris and T. C. T. Yin. A matter of time: internal delays in binaural processing. *TRENDS in Neuroscience*, 30(2):70–78, 2006.

[4] B. Grothe. Sensory systems: New roles for synaptic inhibition in sound localization. *Nat. Reviews Neurosci.*, 4:540–550, 2003.

[5] A. N. Popper and R. R. Fay, editors. *The Mammalian Auditory Pathway*. Springer-Verlag, 1992.

[6] N. B. Cant and R. L. Hyson. Projections from the lateral nucleus of the trapezoid body to the medial superior olivary nucleus in the gerbil. *Hear. Res.*, 58:26–34, 1992.

[7] P. H. Smith, P. X. Joris, and T. C. T Yin. Anatomy and physiology of principal cells of the medial nucleus of the trapezoid body (mntb) of the cat. *J. Neurophysiol.*, 79(6):3127–3142, 1998.

[8] David McAlpine, Dan Jiang, and Alan R. Palmer. A neural code for low-frequency sound localization in mammals. *Nat. Neurosci.*, 4(4), april 2001.

[9] D. J. Tollin. The lateral superior olive: A functional role in sound source localization. *The Neuroscientist*, 9(2):127–143, 2003.

[10] D. W. Grantham. Spatial hearing and related phenomena. In B. J. Moore, editor, *Hearing*, pages 297–345. Academic Press, 1995.

[11] M. J. Goupell and W. M. Hartmann. Interaural fluctuations and the detection of interaural incoherence. ii. brief duration noises. *J. Acoust. Soc. Am.*, 121(4):2127–2136, 2007.

[12] M. Slaney. Auditory toolbox: Version 2, October 1998. Technical Report No. 1998-010. http://rvl4.ecn.purdue.edu/∼malcolm/interval/1998-010/.

[13] W. M. Jenkins and R. B. Masterton. Sound localization: Effects of unilateral lesions in central auditory system. *J. Neurophys.*, 47(6):987–1016, 1982.

[14] R. Y. Litovsky, H. S. Colburn, W. A. Yost, and S. J. Guzman. The precedence effect. *J. Acoust. Soc. Am.*, 106(4):1633–1654, 1999.

[15] T. Hirvonen and V. Pulkki. Interaural coherence estimation with instantaneous ild. In *7th Nordic Signal Processing Symposium*, Reykjavik, Iceland, Jun. 2006. IEEE.

[16] T. Hirvonen and V. Pulkki. Predicting binaural masking level difference and dichotic pitch using instantaneous ild model. In *AES 30th Int. Conf. on Intelligent Audio Environments*, Saariselkä, Finland, March 2007. AES.