Cross-frequency integration in inputs in a functional model of MSO

VILLE PULKKI

Department of Signal Processing and Acoustics, Helsinki University of Technology, POBox 3000, FI-02015 TKK

McAlpine et al. (2001) have measured the response of medial superior olive (MSO) indirectly from inferior colliculus with broadband stimulus having ITD as a parameter. In the resulting ITD functions, it can be seen that the output of MSO is low or random with ITDs corresponding to the best ITD advanced or delayed by a full period at the best frequency, i.e., the sidelobes of the ITD function are relatively low. This effect is striking, since it is known that the impulse response of cochlea is relatively long, meaning that any kind of correlation operation directly between ipsi- and contralateral inputs of MSO cannot produce such low sidelobes. In earlier work of the author a functional model for MSO has been proposed, which follows count-comparison principles. In the present paper it is shown that when it is assumed that the input to each side of a MSO cell is a result of coincidence counting process of a number of adjacent frequency bands of one side, the output of the MSO model has very similar behaviour at the sidelobes when compared with the neurophysiological recordings.

INTRODUCTION

In the previous work, a functional model of binaural decoding based has been proposed containing functional models of the brain organs devoted to binaural cue decoding, namely, the medial superior olive (MSO) and the lateral superior olive (LSO) (Pulkki and Hirvonen, 2007). These organs are modeled using simple linear or nonlinear digital signal processing operations to produce similar statistical input-output relations as have been found in neurophysiological studies (Grothe, 2003; Yin, 2002) with count-comparison principle (Békésy, 1960). The count-comparison principle assumes that the left/right direction is decoded in an organ, which outputs a signal for each hemisphere. The earlier time-of-arrival, or a larger signal level, in one ear canal signal produces larger output to the contralateral hemisphere. This article reviews the earlier-proposed MSO model, and introduces a method for cross-frequency interaction in the input of the model. It is shown, that some cross-frequency interaction is needed to explain the averaged neurophysiological recordings. Also, the effect of incoherent ear canal signals in model response is shown.

AUDITORY PATHWAY

Cochlea and cochlear nucleus

The sound arriving to the inner ear is transformed into neural impulses by the hair cells in the cochlea, which are receptive to certain frequencies more than to others forming auditory filters. Glasberg and Moore used a method to determine the

Proceedings of ISAAR 2009: Binaural Processing and Spatial Hearing. 2nd International Symposium on Auditory and Audiological Research. August 2009, Elsinore, Denmark. Edited by J. M. Buchholz, T. Dau, J. Christensen-Dalsgaard, and T. Poulsen. ISBN: 87-990013-2-2. The Danavox Jubilee Foundation 2010.

properties of the auditory filters; the detection threshold of a signal tone, masked by a noise notched around the signal frequency, is measured as a function of the notch bandwidth (Moore, 1997). The experiments resulted in the Equivalent Rectangular Bandwidth (ERB), which approximates the width of each frequency channel as a function of its center frequency. The auditory filters are often modeled with the gammatone filterbank (Slaney, 1993), which produces a number of channels having one ERB width in frequency.

From the cochlea the signal traverses via the auditory nerve into the cochlear nucleus (CN), which is located in the brainstem. This study is interested in the globular and the spherical bushy cells of CN (GBC and SBC, respectively), which are known to route the neural signals to the organs sensitive to binaural differences (Yin, 2002). At low frequencies, both GBC and SBC phase-lock precisely to specific phases of the ear signals with an accuracy that is greater than that in auditory nerve fibers. At frequencies above about 1 kHz, the accuracy of the phase locking declines fast and there is a notable difference in GBC and SBC responses

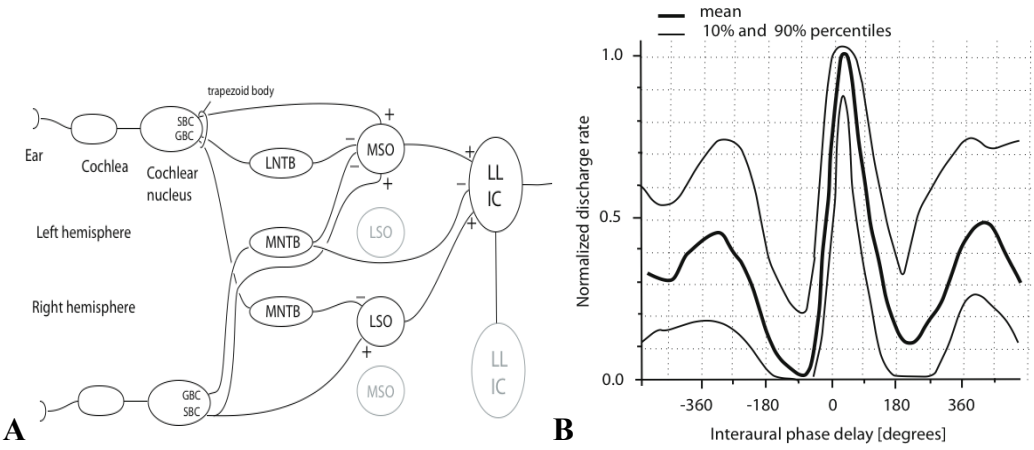

Fig. 1: A) The most important neural connections leading to the left IC found in the mammalian hearing system for binaural decoding below the midbrain. **B)** Averaged output rate of the ITD-sensitive neurons outputs having BFs between 475 and 525 Hz, with white noise stimuli and static interaural phase delay, adapted from (McAlpine *et al.*, 2001).

An important task for GBCs and SBCs is to provide the inputs to the medial and lateral superior olives (MSO and LSO) of both hemispheres through the trapezoid body, as shown in Fig. 1A. The inhibition to MSO and LSO is delivered through the lateral nuclei of the trapezoid body (LNTB) and the medial nuclei of the trapezoid body (MNTB), as shown in the figure, both of which change excitatory signals into inhibitory signals. The MNTB provides an inhibitory input also to the processing level following the level of the superior olives, i.e. to the ventral nuclei of the lateral lemniscus (VNLL). As the VNLL is situated in the pathway leading to the inferior colliculus (IC), the VNLL connection is, for simplicity, drawn to the organ labelled as LL / IC.

Medial superior olive

The MSO receives both excitation and inhibition from both hemispheres, and its cells are known to be sensitive to ITD (Grothe, 2003). The output of the MSO is delivered to the ipsilateral dorsal nucleus of the lateral lemniscus (DNLL) and the IC

(Yin, 2002). The principal cells in the MSO have two dendrites which receive excitatory inputs.

The inferior colliculus (IC) neurons of the guinea pig sensitive to ITD have been measured by McAlpine et al (2001) in tests where the ITD between coherent broad band ear canal signals was a parameter. The responses were measured from a large set of neurons, and the interaural-phase-sensitivity of IC neurons, which shows the maximum response elsewhere than at phase differences ±180°, is assumed to originate from MSO processing. The sensitivity to ±180° is assumed to originate from LSO processing. The mean and approximate 10 % and 90 % percentiles of such responses having BF between 475 and 525 Hz are shown in Fig. 2A as function of phase delay between ears. The responses can be seen to have a repetitive nature with phase delay. The main lobe of the averaged response has the maximum around 45°, which is considered first. The terms 'main lobe' and 'side lobe' in this article refer to the highest peak and the adjacent peaks found in the MSO response, respectively.

It can be seen in the main lobe that when the phase delay is between –90° and 45°, the MSO response in one hemisphere increases with how much the contralateral ear canal signal leads the ipsilateral ear canal signal in phase. It is assumed that this is the most important region in ITD decoding. With other phase delays in the main lobe, namely from 45° to 270°, the MSO response decreases and does not evidently carry information about the sound source position. The increasing and decreasing halves of the main lobe show prominent asymmetry; the response increases with phase delay faster than decreases. The sidelobes correspond to the stimuli where the interaural phase difference corresponds to advance or delay of a full period from 45° phase difference at the BF. They will be discussed in detail below.

The MSO functioning has been described by assuming that the MSO neurons act as coincidence counters (Grothe, 2003), which receive both excitation and inhibition originating from both hemispheres. The contralateral inhibition is assumed to arrive before excitation, which effectively delays the contralateral excitation. Grothe also suggests that, in contrast to contralateral arrival order, ipsilateral inhibition arrives after excitation, which effectively shortens the excitation.

Sidelobes in MSO response

The responses of individual MSO neurons show similar behavior in the main lobe, which can be seen from the fact that the data range is low there, as presented in Fig. 1B. Outside the main lobe the data range is higher. The averaged sidelobes have their maxima with a value of about 0.5 at ITDs corresponding to 45° ± 360°. Such relatively low side lobe heights are assumed to be due to the stimulus waveform. If, for instance, a sinusoid signal was used as stimulus instead of the broad-band noise, the sidelobes would have equal height, as different cycles of sinusoid have the same waveform and are indistinguishable with any measuring device. In this case, the input was broad-band noise, which is assumed to produce such low sidelobe responses.

However, if the MSO is assumed to implement any kind of correlation processing between the ear canal signals, such a low output value when delaying either of the inputs by a single period from 45° is problematic to achieve in computational modeling, if the outputs of single linear auditory filters (e.g. Gammatone filters) are used directly as inputs to the correlator (e.g. multiplication). This follows from the

fact that the impulse responses of the auditory filters include a moderately slowly decaying sinusoid at the BF, which makes subsequent periods in input channels to be inevitably similar.

FUNCTIONAL MODEL

The signals in the model represent the pooled responses of modeled organs. 'Pooling' here means that the average level of activity in each frequency channel is considered rather than the individual responses of the cells.

Peripheral model

The input of the peripheral model is the ear canal signal, and the output is a signal, which models the pooled signal arriving at the trapezoid body. The middle ear is modeled with a filter derived from minimum audible pressure (MAP) recordings. The cochlea is modeled with a linear fourth-order gammatone filter bank (GTFB) with 165 frequency bands spaced with 1/4 ERB. After the filter bank, each frequency channel signal is half-wave rectified.

The phase locking in the cochlear nucleus is implemented by replacing each half-wave by an impulse at a local maximum, the magnitude of which equals to the rms of the half wave. The neural mechanisms of GBCs and SBCs are not modeled differently, as the presented differences at high frequencies are considered minor for this study. In reality, the impulses in the trapezoid body are spread to about 1/10 of the cycle period, as seen in (Joris *et al*, 1994). This spreading is modeled by convolving the impulses with a Gaussian window with similar length; see (Pulkki and Hirvonen 2009) for details. An example output is presented in Fig. 2A.

A time delay compensating the frequency-dependent delay of GTFB is added to each frequency channel. The delay is numerically computed for each frequency band individually, so that the impulse responses of each frequency channel show the highest peaks at the same time instant.

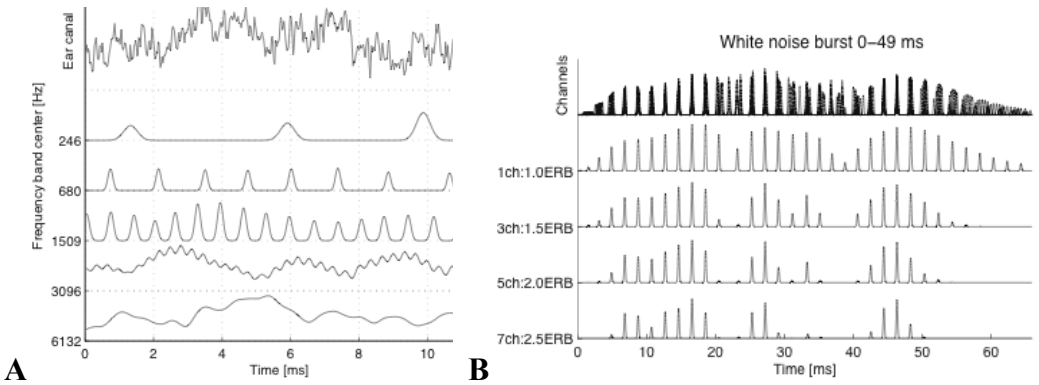

Fig. 2: A) Normalized responses of the periphery model to white noise stimulus at five frequency band. **B)** Effect of geometric averaging of adjacent frequency channels spaced by 1/4 ERBs with white noise stimulus. The 'channels' function show all seven frequency channels from 446 Hz to 565 Hz. The other functions show the result of geometric averaging with a tabulated number of frequency channels and frequency width.

MSO model

In this study, the scope is to imitate the pooled response of the MSO neurons to the pooled responses from the CNs of both hemispheres. The target response for the simulations can be thus set to be the neurophysical results presented in Fig. 1B, as the normalized discharge rate plotted there represents a real pooled MSO response. The model is designed to give an output which is independent of signal level and normalized between zero and one. If the output of the MSO model in one hemisphere is zero, the signal in the ear canal ipsilateral to the MSO model arrives at the same time or earlier than the signal in the ear canal contralateral at the MSO model. If the output is greater than zero, the signal arrives earlier to the contralateral ear. The block diagram of the model is shown in Fig. 3.

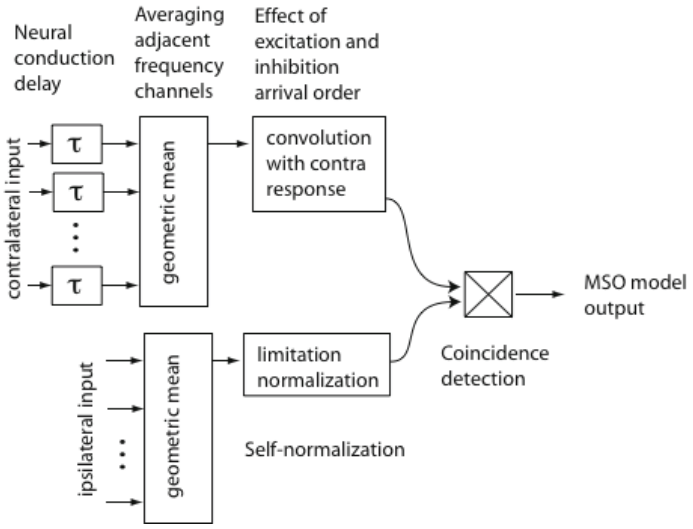

Fig. 3: Block diagram of the MSO model.

MSO inputs

All processing in the MSO model occurring before the binaural interaction is now described. The contralateral input is delayed by a constant amount of time in order to simulate the propagation time difference between the contralateral and the ipsilateral signals. A value has been selected using the knowledge from cat LSO response measurements, where it was found to be about 0.2 ms (Joris, 1996). As it is known that the MSO is nearer the median plane than the LSO, and as the skull diameter of a cat is approximately half of the human skull diameter, it is assumed that the same value of 0.2 ms would roughly model the conduction delay time difference for humans. For the guinea pig, whose skull diameter is 1/4 of the human skull, the delay is set to 0.05 ms. The guinea pig simulation results are shown in Fig. 4.

This model assumes that a number of adjacent frequency bands are combined together with a coincidence counting mechanism, which is modeled here with a running geometric average of the inputs

$$y_j = \sqrt[2W+1]{\prod_{i=j-W}^{j+W} x_i}\,, \qquad \text{(Eq. 1)}$$

where y_j is the result of averaging at the *j*:th frequency channel, *W* is a non-negative parameter which controls the width of the frequency window for averaging, and x_i is the signal of the *i*:th frequency channel from the periphery model. The geometric mean implements a coincidence counting function, or an AND-like function, where the final input is active only when a pulse arrives at the same time from each frequency channel. The selection of the width parameter for averaging is explained below. This section considers only the main lobe of the response, in which the width parameter can be omitted as the main lobe is not affected by the parameter as seen in Fig. 4A.

The ipsilateral input is saturated at 30 dB, and is also scaled in a way that 30 dB saturation level yields the output value of one. It is thus assumed that the level of the MSO output is not independent of the level of the ipsilateral activation insofar as the level is above a constant value. In this model, the level has been selected to equal about 30 dB at 1 kHz. This value was motivated by the fact that the human sensitivity to ITD degrades when the level is reduced under 30 dB as found in ITD-ILD trading tests (Blauert, 1997). Grothe proposed that the ipsilateral inhibition, which is assumed to shorten the effect of excitation in MSO, arrives shortly after excitation. In the presented model, the input arriving to MSO model consists of short Gaussian pulses as shown in Fig. 2A. In principle, the incoming pulses should first be temporally spread to model the effect of ipsilateral excitatory input, and then this spreading should be truncated to model the effect of ipsilateral inhibition. However, it is assumed that such complicated processing is not needed. Instead, no temporal spreading is added in the ipsilateral input, in contrast to it being added to the contralateral input and to the LSO inputs.

The contralateral input is delayed by convolving it with a function *f* that simulates the effect of the inhibition and excitation arrival order. The function has been chosen to be a single period of time-warped raised cosine:

$$f = \left(\cos\left(\left(\Omega\sqrt[4]{n/f_s}\right) - \pi\right) + 1\right)^2 \qquad \text{(Eq. 2)}$$

where Ω is the angular frequency corresponding to the BF of a frequency band, *n* is time index, and f_s is sampling frequency. This gives a good match at 750 Hz frequency band, as seen in Fig. 4A.

Cross-frequency interaction in MSO inputs

The MSO modeling approach assumes that one frequency band in the MSO receives inputs from adjacent frequency channels in the cochlea, and that the AND-like function is computed over the inputs before modeling of cross-hemisphere interaction. This assumption is supported by the fact that a number of synapses are connected to MSO dendrites (Grothe, 2003). The model also assumes that roughly an equal number of adjacent frequency channels are connected to both ipsi- and contralateral inputs in both excitatory and inhibitory inputs.

In practice, this averaging mechanism emphasizes the peaks of envelope of the input signal. It is important to remember here that the delays of the gammatone filters have been compensated at an earlier processing stage. This means that the largest response of an impulse in each frequency channel arrives at the averaging block at the same time, which effectively attenuates the portions of the signal outside the peaks of the waveform. The effect of geometric averaging over frequency channels

is shown in Fig. 2B for white noise. It can be seen that the more of the adjacent channels are used in case of broadband noise, the more the peaks of the broad-band input signal are emphasized.

In the discussion above, the frequency bandwidth of averaging in the MSO input was not discussed. The width of the cross-frequency integration is now selected by fitting the output of the model to neurophysiological results. In the presented model, the gammatone filters having one ERB frequency width are separated by 0.25 ERB from each other, which means that they overlap each other largely. Thus, when a band is added to the frequency window in the averaging, it adds 0.25 ERBs to the total width, which is 1 ERB for a single band. For example, with three consecutive channels the averaging width is 1.5 ERBs. In Fig. 4A, the effect of widening the frequency window in averaging is shown. The best fit with neurophysiological data is obtained when the CN input is temporally averaged over thirteen frequency bands gathering sound from 4 ERB width in the cochlea, as seen in the figure. More neurophysiological and/or psychoacoustical data should be obtained to investigate this phenomenon. It has to be noted, that the Cochlea is modeled using a simple linear filter bank, which neglects nonlinear phenomena. It can be assumed, that when such effects are taken into account better, the width of frequency averaging window fitting best the experimental data changes.

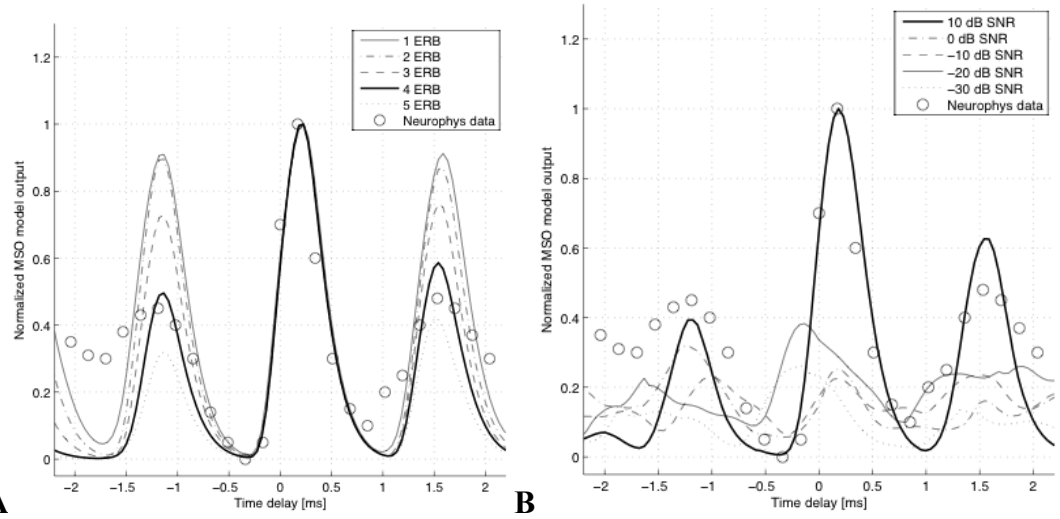

Fig. 4: **A)** Averaged MSO output with coherent white noise input and ITD as a parameter, with different frequency widths in input averaging. **B)** Averaged MSO model output with coherent white noise input with given ITD and tabulated SNR. Additional incoherent white noise has been added to the inputs of the model with tabulated SNRs.

As it can be seen in Fig. 3B, such averaging over frequency channels emphasizes the envelope of the signal, and makes the input pulses to occur sparsely with time. If the ear canals signals are incoherent, the pulses occur at different time instants, and they will not meet in the MSO model, making the MSO model output low. An additional test was conducted to illustrate this. The same input as in the previous test was incorporated, however, incoherent white noise was added to the inputs of the model, with different signal-to-noise ratios (SNR) from -30 dB to 10 dB with 10 dB steps.

The result is shown in Fig. 4B. It can be seen that as the coherence between ear canal decreases, also the MSO response drops with all ITD values. This implies that the MSO model is responsive only when the ear canal signals are coherent on relatively broad frequency region. This seems to be a functionally similar mechanism as has earlier been proposed as an extension to the Jeffress model. In the mechanism only those cues are selected which have been computed from highest cross-correlation peaks (Faller and Merimaa, 2004).

Further tests of the model

The capacity of the model to describe common spatial hearing phenomena has been described in (Pulkki and Hirvonen, 2009), such as lateralization of sines, directional hearing in anechoic space and perception of incoherent ear canal signals and diffuse sound field.

SUMMARY

This article overviews a functional count-comparison MSO model constructed based on recent knowledge on neurophysiology of MSO (Grothe, 2003). It is shown, that when the inputs of the MSO model are geometrically averaged over 4-ERB wide band, a similar results is obtained than in neurophysiological measurements. It is also shown, that with such averaging, the response of the model is high only when the ear canal signals are coherent in the frequency region of the averaging.

ACKNOWLEDGMENT

This research has been funded by the Academy of Finland.

REFERENCES

Blauert, J. (**1997**). Spatial Hearing, revised edition (MIT, Cambridge, MA).

von Békésy, G. (**1960**). Experiments in Hearing (McGraw-Hill, New York).

Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997**). “Modeling auditory processing of amplitude modulation: I. Detection and masking with narrow band carrier,” J. Acoust. Soc. Am., **102**, 2892-2905.

Faller, C, and Merimaa J.: (**2004**) “Source localization in complex listening situations: Selection of binaural cues based on interaural coherence.” J. Acoust. Soc. Am. 116.

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

Joris, P. (**1996**). “Envelope coding in the lateral superior olive. II. Characteristic delays and comparison with responses in the medial superior olive”, J. Neurophys. 76, 2137–2156

Joris, P. X., Carney, L. H., Smith, P. H., and Yin, T. C. (**1994**). “Enhancement of neural synchronization in the anteroventral cochlear nucleus. I. Responses to tones at the characteristic frequency”, J. Neurophysiology 71, 1022–1036.

McAlpine, D., Jiang, D., and Palmer, A. R. (**2001**). “A neural code for low-frequency sound localization in mammals”, Nat. Neurosci. 4.

Moore, B. (1997). *An introduction to the psychology of hearing*, fourth edition (Academic, San Diego).

Pulkki, V. and Hirvonen T., (**2007)** "Computational Count-Comparison Models for ITD and ILD Decoding", ICA 19th Int. Cong. On Acoustics.

Pulkki, V., and Hirvonen T. (**2009**) . "Functional count-comparison model for binaural decoding". *accepted to Acustica united with Acta Acustica.*

Slaney, M. (**1993**). "An effcient implementation of the Patterson-Holdsworth filter bank", Technical Report 35, Apple Computer, Inc.

Yin, T. (**2002**). "Neural mechanisms of encoding binaural localization cues in the auditory brainstem", in *Integrative Functions in the Mammalian Auditory Pathway*, edited by D. Oertel, A. Popper, and R. Fay, 99–159 (Springer, New York).