

Functional rate-code models of the auditory brainstem for predicting lateralization and discrimination data of human binaural perception

Jaroslav Bouse, Václav Vencovský, František Rund,^{a)} and Petr Marsalek^{b)}

Department of Radioelectronics, Faculty of Electrical Engineering, Czech Technical University in Prague, Prague, 166 27, Czech Republic

(Received 22 December 2017; revised 4 December 2018; accepted 5 December 2018; published online 3 January 2019)

This paper presents a rate-code model of binaural interaction inspired by recent neurophysiological findings. The model consists of a peripheral part and a binaural part. The binaural part is composed of models of the medial superior olive (MSO) and the lateral superior olive (LSO), which are parts of the auditory brainstem. The MSO and LSO model outputs are preprocessed in the interaural time difference (ITD) and interaural level difference (ILD) central stages, respectively, which give absolute values of the predicted lateralization at their outputs, allowing a direct comparison with psychophysical data. The predictions obtained with the MSO and LSO models are compared with subjective data on the lateralization of pure tones and narrowband noises, discrimination of the ITD and ILD, and discrimination of the phase warp. The lateralization and discrimination experiments show good agreement with the subjective data. In the case of the phase-warp experiment, the models agree qualitatively with the subjective data. The results demonstrate that rate-code models of MSO and LSO can be used to explain psychophysical data considering lateralization and discrimination based on binaural cues. © 2019 Acoustical Society of America.

<https://doi.org/10.1121/1.5084264>

[GCS]

Pages: 1–15

I. INTRODUCTION

The human auditory system receives the incoming sound through two ears located on opposite sides of the head. Temporal and spectral disparities between the signals in the two ears provide cues about the spatial location of the incoming sound. These cues are the differences in times and levels between the ears. The differences in time are known as interaural time differences (ITDs) and interaural phase differences (IPDs). The differences in level are called interaural level differences (ILDs). ITD and ILD enable the sound to be localized in the horizontal plane. In addition, the body and the pinna reflect the incoming sound, which provides monaural cues, allowing the sound to be localized in the sagittal plane (Blauert, 1997).

According to the duplex theory of sound localization, ITD and ILD are combined to localize a sound in the horizontal plane. The experiments first conducted by Lord Rayleigh lead to tentative idea that for pure tones, ITD would dominate sound localization at low frequencies (originally only below 125 Hz) while ILD would dominate localization at higher frequencies (originally from 500 Hz onwards; Rayleigh, 1907). In free field conditions, pure tones can be localized on the basis of ITD at frequencies up to about 1.5–2 kHz, and can be localized on the basis of ILD from about 1.5–4 kHz (Mills, 1960; Stevens and Newman, 1936). When noise bands were used, it was found that ITD

dominates localization at low frequencies up to the boundary estimated at between 1.5 and 2.5 kHz (Blauert, 1997; Wightman and Kistler, 1992). A recent study by Hartmann *et al.* (2016) showed that the natural salience of ILD in a free field may be high even at frequencies below 1 kHz.

For pure tones, the smallest detectable change in ITD depends largely on the frequency. While the threshold can be as low as 10 μ s at 0.8 kHz, it increases rapidly when the frequency is increased up to 1.2 kHz or is reduced to 0.2 kHz (Brughera *et al.*, 2013; Klumpp and Eady, 1956; Zwillocki and Feldman, 1956). In contrast, the ILD threshold is almost constant (0.5 dB) across frequencies between 0.5 and 8 kHz with a peak at 1 kHz (Grantham, 1984; Mills, 1960; Yost and Dye, 1988).

Listening through headphones often results in an auditory sensation localized within the head. The lateral position of this sensation is known as the lateralization (Blauert, 1997; Moore, 2003). The lateralization is moved toward one of the ears by manipulating the ITD, ILD, or both (Sayers, 1964; Yost, 1981).

If the tones in the two ears have slightly different frequencies, it may cause binaural beats. The perceived position of the sound oscillates between the left and right ear for a frequency difference between the tones of about 2 Hz (Moore, 2003). Siveke *et al.* (2008) introduced a broadband binaural beat stimulus called phase warp that can be created by an up/down circular shift of the phase spectrum of the noise in one channel by a defined “beat” frequency. When such a sound is listened to through headphones, the phase-warp stimulus produces the sensation of a sound source rotating around the listener’s head. The frequency of the

^{a)}Electronic mail: xrund@fel.cvut.cz

^{b)}Also at: Institute of Pathological Physiology, First Medical Faculty, Charles University, U Nemocnice 5, Praha 2, 128 53, Czech Republic.

rotation is given by the beat frequency. The sensation of rotation disappears when the beat frequency exceeds 10 Hz. The sensation then changes to roughness.

In mammals, ITD and ILD are decoded in the medial superior olive (MSO) and the lateral superior olive (LSO), respectively (Grothe, 2003; Grothe *et al.*, 2010; Joris, 1996; Joris and Yin, 1995; Marquardt and McAlpine, 2007; McAlpine *et al.*, 2001; Tollin, 2003; Tollin and Yin, 2005). While MSO is sensitive to ITD, LSO is sensitive to both ITD and ILD with an emphasis on ILD (Joris, 1996; Joris and Yin, 1995; Tollin, 2003; Tollin and Yin, 2005).

A possible ITD detection mechanism was proposed by Jeffress (1948). He suggested that neural discharges coming from the left and right ears propagate through delay lines to coincidence detectors. The detectors fire if the discharges from the left and right ears arrive within a short time window, i.e., when the input delay line connections effectively cancel out the ITD. The theory assumes that each neuron is tuned to a specific ITD, and its activity is greatest for this ITD. This mechanism can be simulated using interaural cross-correlation (Cherry and Sayers, 1956; Stern and Colburn, 1978). The Jeffress delay line mechanism was subsequently extended to account for the detection of ILD (Breebaart *et al.*, 2001; Gaik, 1993).

Recently, the “delay line” mechanism of Jeffress has been questioned in some neurophysiological studies, which have shown that neurons in the mammalian MSO respond maximally for ITD corresponding to a 45-deg IPD (0.125 of the cycle; Grothe, 2003; McAlpine and Grothe, 2003). These studies propose that the difference in the relative spike rate between MSO neurons in the left and right sides of the brain encodes the spatial direction of the sound. In humans, the magneto-encephalography study by Salminen *et al.* (2010) gave evidence for such a “hemifield rate-code” of auditory space.

Models based on the hemifield rate-code are sometimes referred to as “count comparison” (Encke and Hemmert, 2018; Pulkki and Hirvonen, 2009; Takanen *et al.*, 2014). This term was originally introduced by Colburn (1978) to describe the principle of the model designed by von Békésy (1930). The von Békésy model assumes that each ear innervates the same neurons, and that each neuron may be tuned “left” if the signal from the left ear arrives shortly before the signal from the right ear, or tuned “right” in the opposite case. The firing activity of neurons tuned left and right is then compared to give the direction of the sound source. Since the number of excited neurons increases with a rising level of the sound, this model accounts for the time-intensity trading. van Bergeijk (1962) later adjusted the model to account for the fact that MSO and LSO are composed of paired units placed symmetrically relative to the median plane.

One of the first rate-code models based on neurophysiological data is the functional model of the MSO and LSO designed by Pulkki and Hirvonen (2009). This model uses signal processing to reproduce data from neurophysiological studies. Takanen *et al.* (2014) extended the model by integrating the MSO and LSO outputs into a visual map, which allowed a direct comparison to be made between the

listening test results and model predictions. In addition, the model was extended with a hypothetical wideband MSO, which would account for the detection of ITD in the envelope of high-frequency sounds.

Dietz *et al.* (2008) proposed in their model that the human auditory brainstem can effectively encode and decode the IPDs from the interaural transfer function (ITF; Blauert, 1997). By filtering the peripheral model output using two parallel complex bandpass filters, Dietz obtained two ITFs, which corresponded to the fine structure and the envelope of the IPD. The firing rates of the units simulating left and right MSOs are then calculated. When the fine structure and envelope of the ITF are used, the model can account for binaural masking level differences (BMLDs) and lateralization data. It has been also proven that it could simulate experimental data on broadband binaural beats (phase warp; Siveke *et al.*, 2008). The model was further improved by adding a module for calculating the lateralization based on ILD. The improved model was utilized as a front-end for a direction estimate of concurrent speakers in a binaural signal (Dietz *et al.*, 2011). Dietz *et al.* (2009) also extended the original IPD model with a quantitative estimate of the perceived lateralization, which successfully simulated higher auditory processing. The new model combines the temporal lateralization cues from the fine structure and the envelope of the binaural input signal, giving a single value for the overall perceived lateralization.

In addition to the phenomenological rate-code models mentioned above, Encke and Hemmert (2018) presented their physiologically plausible, spiking neuron network model of the mammalian MSO. The authors used two methods to decode spatial information. The first method—the linear opponent decoder—was able to mimic the ITD threshold data, but its predictions were dependent on the overall sound pressure level (SPL) of the stimulus. The second method was based on a simple artificial neural network (ANN) with inputs from the spiking outputs of MSO and auditory nerve fiber models. With ANN, the model was able to predict static ITD imposed on the amplitude modulated tone and speech stimuli. It was also able to track the transient change of ITD for a sine sweep stimulus.

The Jeffress family of models has been used successfully to predict human psychoacoustical data, but from a physiological point of view the existence of such a neural circuit in the mammalian brain is questionable. However, there is evidence for a rate-code in the mammalian brain, though the presented rate-code models have not explained the psychoacoustical data in a quantitative manner. The aim of this paper is to reduce the gap between neurophysiology and psychophysics by introducing functional rate-code binaural models of MSO and LSO, which will meet three main criteria: (1) they will be based on current neurophysiological findings, (2) their output will give a quantitative representation of subjective lateralization based on ITD or ILD at a corresponding frequency, and (3) they will have low structural complexity and also low computational complexity for low power applications (Vítek *et al.*, 2011). The model designed by Pulkki and Hirvonen (2009) satisfies points (1) and (3), but gives only the relative firing rate at the model output. Enhancing the model by a visual map (Takanen

et al., 2014) solves this problem, but significantly increases the overall structural complexity of the model. The IPD models of *Dietz et al.* (2009, 2011) and *Dietz et al.* (2008) fulfill all three criteria, but calculating IPDs and ILDs from the ITF might be considered artificial. The model of *Encke and Hemmert* (2018) satisfies criterion (1) perfectly, as it is a physiological model, which also limits the model in computational complexity [criterion (3)]. The model also accurately predicts human discrimination data with pure tones with ITD, and in other experiments it predicts ITD. However, there is no mapping of this data to subjective lateralization. For the reasons mentioned here, we propose new functional rate-code models of the human auditory brainstem, which will be presented below.

II. A PROPOSED MODEL OF BINAURAL INTERACTION

The proposed model of binaural interaction is composed of two main parts: peripheral and binaural (see Fig. 1). The peripheral part of the model simulates the function of the auditory periphery. It utilizes algorithms adapted from the Auditory Modeling (AM) Toolbox (*Søndergaard and Majdak*, 2013). The binaural part consists of original models of MSO and LSO connected to the corresponding ITD and ILD central stages. A simpler version of the model was presented in *Bouse and Vencovsky* (2015). MATLAB source codes of the model are available online.¹

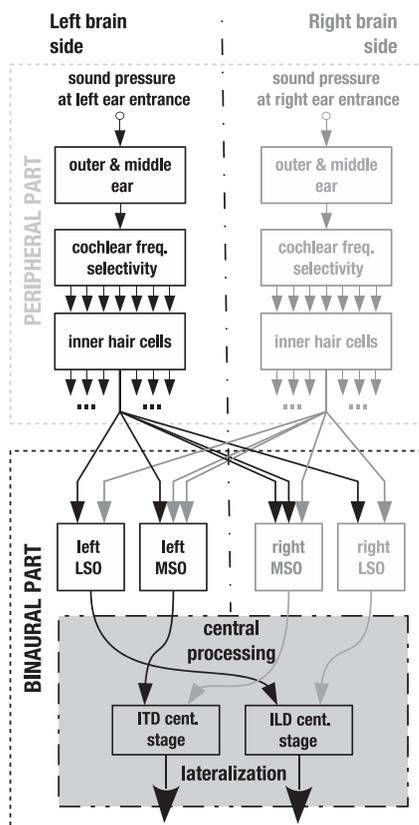


FIG. 1. Schematic diagram of models of MSO and LSO. The diagram is divided vertically by a dashed-dotted line to account for the left and right sides of the brain, and horizontally by dotted boxes to split the models into the peripheral and binaural parts. The gray part highlights processing in the higher stages of the brain.

A. Peripheral part of the model

The peripheral part models the transformation of the incoming sound into the average response of a population of neurons tuned to the specific characteristic frequency in the auditory nerve fiber. This part consists of three functional blocks: the outer ear and middle ear, cochlear frequency selectivity, and the inner hair cells.

The frequency response of the outer ear (headphone to eardrum pressure; *Pralong and Carlile*, 1996) is modeled by a 512th-order finite impulse response (FIR) filter. The transfer function of the middle ear (*Goode et al.*, 1994) is then modeled by an FIR filter of the same order. The model does not account for the middle ear reflex.

The cochlear selectivity is modeled by a dual resonance nonlinear (DRNL) filter bank (*Lopez-Poveda and Meddis*, 2001), which divides the input signal into 70 peripheral channels. The frequency spacing of the auditory filters in the bank was set to be equal to one-half of the equivalent rectangular bandwidth (ERB; *Moore and Glasberg*, 1983), which is calculated as follows:

$$\text{ERB}(f_c) = 24.7 + 0.108f_c, \quad (1)$$

where f_c corresponds to the central frequency (CF) of the peripheral filter in Hz. For all simulations, the CFs were set to be in the range between 0.1 and 14 kHz. In each auditory filter of the bank, the signal propagates through two independent processing pathways: linear and nonlinear. The two paths are joined at the output, and the signals from the two paths are added together. The nonlinear processing path dominates the output at low signal levels and decays with increasing signal level until the output becomes mostly dominated by the linear processing path (*Lopez-Poveda and Meddis*, 2001). We chose this type of model since it accounts for the compressive nonlinearity observed in the input/output functions of the basilar membrane response.

The block simulating the mechano-electrical transduction by the inner hair cell and the auditory nerve fiber complex consists of a half-wave rectifier followed by a low-pass filter (LPF). While the half-wave rectification accounts for the actual mechano-electrical transduction, the low-pass filtering simulates the loss of phase locking of the neuronal signal to the fine structure of the incoming wave for frequencies above 1.5 kHz (*Bernstein and Trahiotis*, 1985; *Weiss and Rose*, 1988). In the present study, a LPF with the same parameters as in studies by *Breebaart et al.* (2001) and *Dietz et al.* (2008) is utilized, i.e., a fifth-order Butterworth infinite impulse response (IIR) filter with a cut-off frequency of 760 Hz.

B. Binaural part of the model

The binaural part consists of two separate original computational models mimicking the behavior of the MSOs and LSOs. Both models follow the rate-code principle (*Colburn*, 1978; *Pulkki and Hirvonen*, 2009; *Takanen et al.*, 2014; *von Békésy*, 1930). There are separate MSO and LSO models for each side of the brain. Each model receives information from the left and right peripheries. The lateralization based

on IPD/ ITD (MSO) or ILD (LSO) is calculated in the corresponding central stages, where a comparison is made between the activities in the models for both sides of the brain. In the present study, the outputs of MSO and LSO are not combined like they are in the auditory pathway (Grothe *et al.*, 2010).

The internal processing of the models is symmetrical for both sides of the brain. Therefore, the signal coming from the periphery at the same side as analyzed in the MSO/LSO model will be referred to as ipsilateral, and the signal from the opposite side will be referred to as contralateral.

1. Model of the MSO

A schematic diagram of the MSO models for both sides of the brain (black line = left, gray line = right) is depicted in Fig. 2. The MSO model is an excitation-inhibition type with two excitatory inputs from both the ipsilateral and contralateral sound peripheries, and one inhibitory input from the contralateral sound periphery. Each input is first processed by a LPF; the third-order low-pass Butterworth filter with a cut-off frequency $f_{\text{cut}} = 1.1$ kHz and gain $G(f) = 1/\sqrt{(f/f_{\text{cut}} + f^6)}$ then reduces the sensitivity of the model sensitivity to ITDs for frequencies higher than about 1.1 kHz. Although this filter has a higher cutoff than the preceding filter in the peripheral part, its function is important to account for the worse ITD discrimination at frequencies >1 kHz.

The physiology data from bats' MSO (Grothe, 1994) and gerbils' MSO (Brand *et al.*, 2002; Roberts *et al.*, 2013) indicate that contralateral inhibition can precede ipsilateral excitation. A small constant delay τ_{MS} of 0.3 ms is inserted into each excitatory signal. In contrast, the excitatory inputs to the MSO from both peripheries have very similar overall conductance delays (Grothe, 2003). Therefore, there is no additional time lag between the ipsilateral and contralateral excitation signals in the MSO model.

The calculation block is the first binaural relay in the model. In the first step, it subtracts the preprocessed contralateral inhibition input I_c from the contralateral excitation input E_c . Then it calculates the mutual signal power of the

product and ipsilateral excitation signal E_i . The output of the calculation block is given by

$$\text{Calc}[n] = E_i[n](E_c[n] - I_c[n]), \quad (2)$$

where n is the sample number. Since a neural signal cannot be negative, the output of the calculation block is half-wave rectified, and this signal is then denoted with subscript h .

The calculation block output contains a number of sharp peaks followed by troughs, which can cause transient variation of the MSO model output even for stationary input signals. Therefore, an envelope is calculated from the half-wave rectified MSO signal in a self-weighted moving average block (Pulkki and Hirvonen, 2009). The envelope is calculated using the formula

$$\text{MSO}_{\text{left or r}}[n] = \frac{\text{Calc}_h^3[n] * H_1[n]}{\text{Calc}_h^2[n] * H_1[n]}, \quad (3)$$

for

$$H_1[n] = (1 - \exp(-1/(f_s \tau_1))) \exp(n - 1/(f_s \tau_1)), \quad (4)$$

where $\text{Calc}_h[n]$ is the half-wave rectified output of the calculation block, $H_1[n]$ is the impulse response of the first-order IIR filter, τ_1 is the time constant of the filter, f_s is the sampling frequency, and “*” denotes convolution. The time constant of the filter τ_1 (2.5 ms) corresponds to the 64 Hz cut-off frequency, which simulates the relative decrease in temporal binaural resolution on higher frequencies reported by Siveke *et al.* (2008).

a. ITD central stage. In this paper, we combine outputs from both sides of the brain in an ITD central stage. This stage is phenomenologically motivated to transform the MSO model outputs into the subjective lateralization scale. In the first place, we have to consider how the outputs of the left and right MSO models vary in relation to the IPD. The right MSO model output is at a maximum when the signal at the left ear has an IPD of minus 50 deg (0.15 of the cycle); at

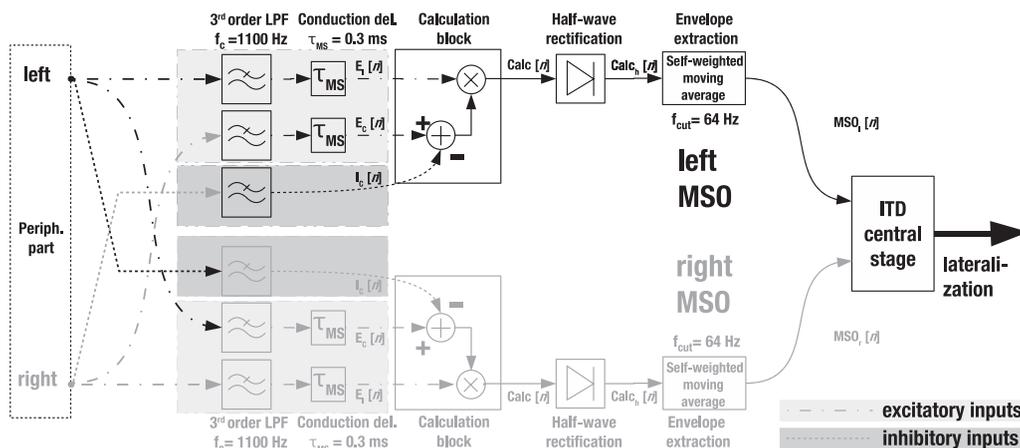


FIG. 2. Schematic diagram of the MSO models for the left (black connection) and right (gray connection) peripheries with the ITD central stage. Each MSO model has three inputs, two excitatory inputs (light gray background) from both peripheries, and one inhibitory input (dark gray background) from the contralateral periphery.

the same moment in time, the left MSO output is minimal and vice versa for a 50-deg IPD. This behavior is consistent with the physiology of the mammalian MSO, where the maximal firing rate is observed around 45 deg (0.125 cycles), when the contralateral ear is leading (Grothe, 2003). The maximum occurs around the 50-deg IPD, independently from the input signal frequency [see Fig. 3(A)]. The MSO model behaves in this way due to the algorithm that is used. The algorithm multiplies the ipsilateral and contralateral inputs and then subtracts the result by the 0.3 ms shifted contralateral input multiplied with the ipsilateral input. However, this processing gives the maximum MSO response at about the same IPD if the peripheral ear model contains half-wave rectification and LPFs limiting the phase locking. The LP filtering shapes the fine structure of the half-wave rectified signal in a way that then produces the desired result. The MSO model also shows similar behavior to that measured on neurons in the guinea pig’s inferior colliculus neurons (McAlpine *et al.*, 2001), i.e., the broadest ITD function for low CFs with peaks at high ITDs and sharper tuning for high CFs with peaks at low ITDs [see Fig. 3(B)]. For both CFs, the predicted responses have a close to zero ITD shallower slope than the physiological data (McAlpine *et al.*, 2001).

The MSO outputs are half-wave rectified and, if one of the MSO models has a zero output, the output of the other MSO model is also set to zero. The two ratios between the left and right MSO models are then computed as follows:

$$r_R[n] = \frac{MSO_l[n]}{MSO_r[n]}, \quad r_L[n] = \frac{MSO_r[n]}{MSO_l[n]}, \quad (5)$$

where MSO_l is a signal from the left MSO model, and MSO_r is a signal from the right MSO model. Ratio r_R is inverse to r_L ; this ensures that, for non-zero MSO model outputs, one of the ratios will always lie between zero and one. If the sound is perceived on the left, the ratio with subscript “L” will be larger than the ratio with subscript “R” symmetrically for the sound perceived on the right. Nevertheless, the larger ratio can be used as a bias between the left or right side, and the smaller ratio is better suited for evaluating lateralization because it is bounded between zero and almost one. The lateralization map ranging from minus one (left ear) to plus one (right ear) is obtained by subtracting unity from r_L if the

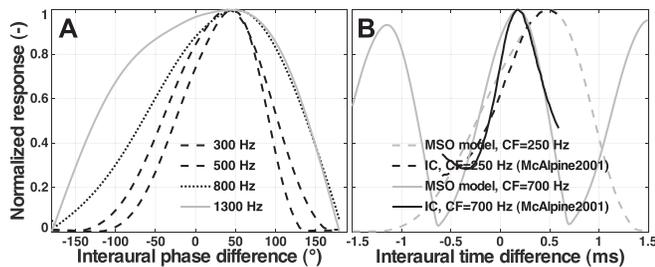


FIG. 3. (A) The normalized responses of left MSO model to the pure tones of varying frequency and IPD. (B) The normalized responses for two CFs (250 and 700 Hz) of left MSO (gray line) with additional side-lobes and pooled responses from a guinea pig’s inferior colliculus (black line), reproduced from McAlpine *et al.* (2001), to interaurally delayed broadband noise.

lateralization is “right-sided” and by subtracting r_R from unity if it is “left-sided”

$$L_{MSO}[n] = \text{sgn}(r_L[n] - r_R[n]) \times (\min(r_R[n], r_L[n]) - 1) + u[n], \quad (6)$$

where L_{MSO} is the lateralization predicted by the MSO model, $u[n]$ is the MSO internal noise, “sgn” is the signum operator, and “min” is the minimum operator. The overall sensitivity of the model to the ITDs is limited only by the mathematical operations within the model. Thus, in order to match this sensitivity with human psychoacoustical data, the internal Gaussian noise $u[n]$ is added into all channels of the calculated lateralization. The variation of this noise was set experimentally in order to match model performance with the ITD discrimination data of Brughera *et al.* (2013) at 1 kHz, and it is constant for all the channels.

2. Model of the LSO

Figure 4 is a schematic diagram of the LSO models for both brain sides (black lines = left, gray lines = right). The LSO model is an excitation-inhibition type with the excitatory input from the ipsilateral periphery and the inhibitory input from the contralateral periphery. Human sensitivity to the intensity disparities between the two ears has repeatedly been reported as logarithmically dependent (Moore, 2003). The inputs of the LSO models are therefore first compressed by a power of 0.24. The compression has similar effect as the logarithm of the ipsilateral and contralateral signals. This compression is used even though our cochlear model also contains compressive nonlinearity. This is similar to the LSO model in Takanen *et al.* (2014). On the basis of data from cats’ LSO (Joris, 1996), the contralateral inhibitory signal is delayed by τ_{LS} equal to 0.2 ms. The function of the mammalian LSO can be characterized as a fast subtraction unit between contralateral and ipsilateral neural signals (Bures, 2012; Bures and Marsalek, 2013; Joris and Yin, 1995). The relative speed of the system is simulated by a first-order IIR filter with a time constant of 0.1 ms. The process of subtraction is then simulated in a subtraction block, where the contralateral inhibitory signal is first subtracted from the ipsilateral inhibitory signal, sample by sample. The product is amplified by a gain $A = 100$ and is limited between -1 and 1 by a hyperbolic tangent function. The amplification gain, together with the limitation, successfully simulates the limit of the maximum firing rate of LSO cells, which occurs around 18 dB ILD (Tollin and Yin, 2005).

The processing of the subtraction block is given by

$$\text{Sub}[n] = \tanh(A(E_i[n] - I_c[n])), \quad (7)$$

where “Sub” is the output of the LSO subtraction unit, “tanh” is the hyperbolic tangent function, E_i is the excitation from the ipsilateral periphery, I_c is the inhibition from the contralateral periphery, and A is the linear gain.

All negative samples after the subtraction block are zeroed in the half-wave rectification block. The rectification, together with the subtraction, produces ripples at the output,

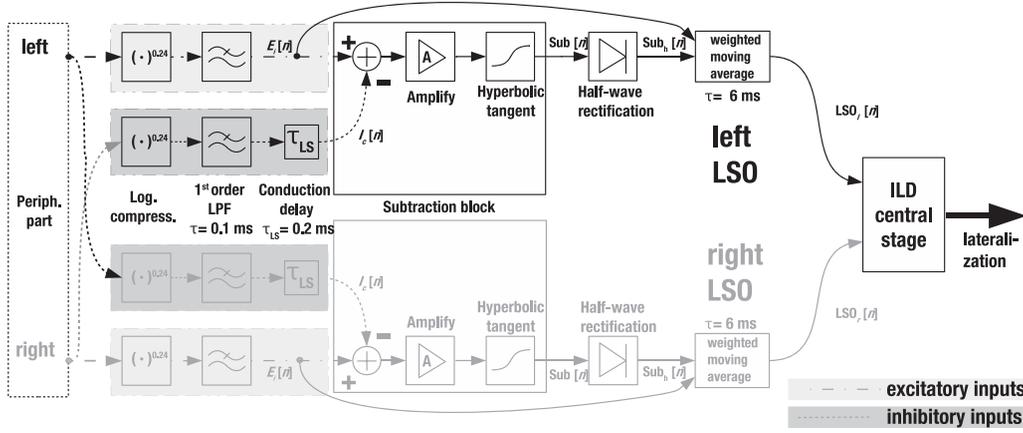


FIG. 4. Schematic diagram of LSO models for the left (black connection) and right (gray connection) sides of the brain with ILD central stage. The LSO model on one side has two inputs, an excitatory input (light gray background) from the ipsilateral periphery and an inhibitory input (dark gray background) from the contralateral periphery.

which are smoothed by the weighted moving average unit. The design of this unit originates from Pulkki and Hirvonen (2009), and acts as a sample and hold RC (resistor-capacitor) circuit with a time constant τ_2 of 6 ms. This processing is not inspired by neurophysiology, but it helps to produce the desired decrease in sensitivity of the model around 1 kHz. The LSO output after the weighted moving average unit is given by

$$\text{LSO}_{\text{left}}[n] = \frac{(\text{Sub}_h[n]E_i^2[n]) * H_2[n]}{E_i^2[n] * H_2[n]}, \quad (8)$$

for

$$H_2[n] = (1 - \exp(-1/(f_s\tau_2))) \exp(n - 1/(f_s\tau_2)), \quad (9)$$

where Sub_h is the half-wave rectified output of the LSO subtraction unit, E_i is the excitation from the ipsilateral periphery, and “*” denotes convolution.

a. ILD central stage. The output of each LSO unit is proportional to the perceived lateral displacement of the sound source on the corresponding side of the brain. In the case of low-frequency signals, there is relatively high activity in both LSO units for near-zero ILDs. This activity is caused by the delay of the inhibitory signal from the contralateral brain side. Therefore, a simple central stage calculates the lateralization based on the ILD by

$$L_{\text{LSO}}[n] = \text{LSO}_r[n] - \text{LSO}_l[n] + v[n], \quad (10)$$

where LSO_r represents the signal from the right LSO unit, and LSO_l represents the signal from the left LSO unit, and $v[n]$ represents the LSO internal noise. Here, sensitivity to the ILD of the LSO model, similarly to the MSO model, is limited only by the internal mathematical operations, which are more than one order lower than human psychoacoustical data. We therefore add Gaussian noise into every channel of the calculated lateralization. The variation of the noise was chosen to match human sensitivity to changes in ILD at 1 kHz (Yost and Dye, 1988), and it is constant for all channels. The ILD central stage is phenomenologically motivated, though such simple processing is more plausibly

presented physiologically than the ITD central stage. The predicted lateralization ranges in the interval between -1 and 1 , where “ -1 ” stands for perception near the left ear, “ 0 ” for perception near the center of the head, and “ 1 ” for perception near the right ear.

III. EXPERIMENTAL METHODS

The experimental method, the stimulus details, and the simulation parameters are described below. In the present study, an experiment was conducted to obtain the lateralization of narrowband noise (NBN) with IPD or ILD. The data on lateralization of pure tones with IPD or ILD, discrimination of ITD, discrimination of ILD, and discrimination of phase warp were taken from the literature (Brughera *et al.*, 2013; Dietz *et al.*, 2008; Yost, 1981; Yost and Dye, 1988).

A. Subjects

Seven subjects (including one female) participated in the NBN with IPD experiment, and eight subjects (including two females) participated in the NBN with ILD experiment. The participants were aged between 20 and 46 yr (all except for one were below 40 yr of age). One subject participated in both experiments. Subjects had no or little prior experience with this type of listening test. Their pure-tone hearing thresholds were within a range of 15 dB hearing level for frequencies between 0.25 and 8 kHz (normal hearing). The subjects took part in the experiments voluntarily, and all procedures were in accordance with current Acoustical Society of America (ASA) ethical principles.

B. Apparatus and procedure

The NBN listening experiment took place in a sound-insulated booth. The subjects were seated in front of a computer monitor and listened to the stimuli through headphones (in the IPD experiment, Sennheiser HD 595, Wedemark, Germany; in the ILD experiment, Sennheiser HD 650) connected to the sound card output (RME Fireface UC, Haimhausen, Germany). The headphones were calibrated to maintain the same SPL in both channels of the test material.

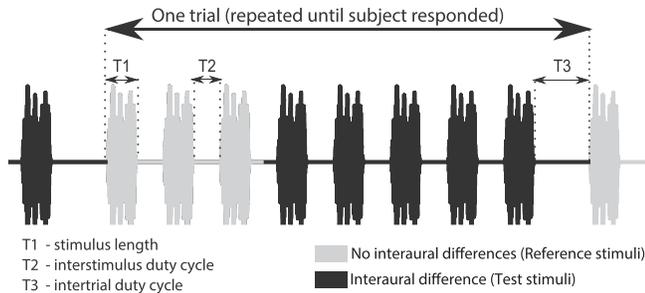


FIG. 5. Time diagram of one stimuli train. The train was composed of three reference NBNs without interaural differences, followed by five stimuli with the same testing interaural difference. All eight stimuli were pulsed with a 50% duty cycle. The stimuli train was repeated after 700 ms of silence until the listener responded.

The experimental procedure was similar to that used by Yost (1981). The listeners indicated a perceived lateral position of the sound on a graphic scale. The scale was represented as a movable pointer on a drawing of a head. The position of the pointer represented the perceived lateralization of the sound inside the head. Each slider position was linearly mapped to a numerical value ranging from -10 to 10 , where “ -10 ” corresponded to the maximum lateral displacement to the left, and “ 10 ” corresponded to the maximum lateral displacement to the right. The experiment was organized in four listening sessions, two for each CF (350 and 760 Hz). Listeners were required to take a minimum break of 5 min after each session. All stimuli were presented five times to a listener during a given session in a random order. From each of the listening sessions we obtained ten complete sets of subjective data per CF. In the first listening session for each CF, the first two sets were discarded to provide enough time for the listeners to adjust to the procedure. In the second session for each CF, the first set of data was discarded. Overall, seven sets of subjective data for each CF were used in the evaluation. The stimuli were presented to the listener in the form of a repetitive stimuli train (see Fig. 5).

C. Stimuli

1. Lateralization of pure tones with IPD or ILD

The data on lateralization of pure tones with IPD or ILD were taken from Yost (1981). The pure tones had frequencies of 0.2, 0.5, 0.75, 1, and 1.5 kHz for the experiment with IPDs, and 0.2, 0.5, 1, 2, and 5 kHz for the experiment with ILDs. They were 100-ms long with 8-ms-long on- and off-raised-cosine ramps. The sound level was 50 dB relative to the hearing level according to ANSI S3.6-1996 (1996). In the IPD experiments, the IPDs were varied from -150 to 180 deg with a 30-deg step, while the ILDs were varied from -18 dB to 18 dB with a 3-dB step in the ILD experiments.

2. Lateralization of NBN with IPD or ILD

NBN with a bandwidth equal to 1 ERB [see Eq. (1)] was used as a stimulus in the experiment and the simulation. Two distinct CFs f_c (350 and 760 Hz) of NBN were used. NBNs were generated in each listening session in the

frequency domain with random amplitude and phase in the passband frequencies. The stimuli were 100-ms long with an 8-ms-long on- and off-raised-cosine ramps. The level was set to 50 dB relative to the subject’s hearing level at the CF f_c . In the IPD experiment, the IPDs were varied from -150 to 180 deg with steps of 30 deg. The IPD between the left and right ear channels was created in the frequency domain: the desired IPD was imposed onto the phase spectrum of the to-be-delayed signal, after which the signal was transformed back to the time domain. In the ILD experiment, the ILDs were varied within the range from the interval from -18 to 18 dB with steps of 3 dB with -20 and 20 dB ILD in addition. The ILD was imposed on the stimuli by amplifying one channel by $ILD/2$ and attenuating the other channel by the same amount.

3. Discrimination of ITD

The ITD discrimination threshold of the MSO model was obtained using pure tones of the same parameters as in the experiment conducted by Brughera *et al.* (2013). The pure tones were 500-ms long (100-ms-long on- and off-raised-cosine ramps) with frequencies of 0.25, 0.5, 0.7, 0.8, 0.9, 1, 1.2, 1.25, 1.3, and 1.35 kHz, and were presented at 70 dB SPL.

4. Discrimination of ILD

Pure tones with the same parameters as in Yost and Dye (1988) were used to obtain the ILD discrimination threshold of the LSO model. The 250-ms-long pure tones (10-ms-long on- and off-raised cosine ramps) were presented at a nominal SPL of 60 dB with frequencies of 0.2, 0.5, 1, 2, and 5 kHz.

5. Discrimination of phase warp

In the simulation of the discrimination of the phase warp, the parameters of the stimuli were as in the experiment conducted by Dietz *et al.* (2008). The phase-warp stimuli were 1-s long (20-ms-long on- and off raised-cosine ramps) and had a 65 dB SPL. The left channel of the stimulus was created in the frequency domain from noise with a constant amplitude and random phase spectra (normally distributed) in the passband. The passband of the noise ranged from 10 Hz to either 550 Hz or 1100 Hz. The right channel was created by a cyclical-frequency shift of the phase spectrum of the left channel by beat frequency f_b of the phase-warp stimulus, i.e., the whole phase spectrum was shifted in frequency by f_b , and the part of the spectra that was above the passband of the phase warp was moved to the start of the passband. In order to keep to the procedure of the original experiment, binaurally uncorrelated noise with a passband from 550 Hz or 1100 Hz to 48 kHz of the same spectral level was added to the phase-warp stimulus. The phase-warp beat frequency and its bandwidth were the variables in the first part of the experiment.

In the second part of the experiment, the phase warp $p[n]$ was mixed with binaurally uncorrelated Gaussian noise $w[n]$, which resulted in a mixed signal $s[n]$, given by

$$s[n] = p[n]r + w[n](1 - r), \quad (11)$$

where ratio r is calculated from modulation depth m by

$$r = \frac{1}{1 + \sqrt{1/m - 1}}. \quad (12)$$

The modulation depth m was the only variable in the second part of the experiment. The dB value of modulation depth is calculated as $10 \log_{10}(m)$.

D. Simulation procedure

These model results were obtained for the stationary part of the output signals only. The input transient was omitted. Therefore, the model is not intended to be used for studying the onset dominance of human sound localization or the precedence effect.

1. Lateralization of pure tones with IPD or ILD

In the simulation of pure-tone lateralization, the mean values were calculated from the stationary part of the ITD or ILD central stage responses. Only a single band with CF nearest to the pure-tone frequency was taken into account. In the IPD experiment, only the ITD central stage response was taken into account. By contrast, in the ILD experiment, only the ILD central stage was taken into account.

2. Lateralization of NBN with IPD or ILD

During the NBN noise simulation, only one stimulus with the interaural difference from the stimuli train (see Fig. 5) was analyzed using the MSO or LSO model. Only a single band with CF nearest to the narrowband CF was taken into account. Afterward, the same simulation procedure was used as had been used in the lateralization of pure tones with IPD or ILD.

3. Discrimination of ITD

In the ITD discrimination simulation, we assumed that the human auditory system could save the “pattern” of one stimulus and compare it with the pattern of another stimulus. In this case, discrimination index d' (Sakitt, 1973) is considered as the ideal observer performance

$$d'(A, B) = \frac{|\mu_A - \mu_B|}{\sqrt{\sigma_A \sigma_B}}, \quad (13)$$

where μ_A , μ_B , σ_A , and σ_B are the means and the standard deviations, respectively, of the output of the MSO model ITD central stage for stimulus A and stimulus B. The discrimination index is calculated only for a single band, the CF of which is nearest to the CF of the pure tone. The observer judgments were based on the difference in ITD between stimulus A and stimulus B; this difference will henceforth be denoted as ΔITD . Stimulus A had ITD equal to $-\Delta\text{ITD}/2$ (lateralized to the left ear), while stimulus B had ITD equal to $\Delta\text{ITD}/2$ (lateralized to the right ear). The discriminable ITD was detected when discrimination index d' exceeded the threshold limit and, at the same time, the model predicted the correct left or right lateralization shift between

the tone in the first and second intervals. The threshold limit (1.14) was chosen to estimate the 79.4% correct point of the psychometric function according to Hacker and Ratcliff (1979). The same point of the psychometric function was targeted in the original listening experiment (Brughera *et al.*, 2013) using a three-down, one-up adaptive staircase procedure (Levitt, 1971). The simulation followed the two-interval, two-alternative forced-choice (2AFC) paradigm as in the listening experiment conducted by Brughera *et al.* (2013). The ΔITD started at $100 \mu\text{s}$ and was decreased with steps of $17 \mu\text{s}$ until four reversals were reached, after which the step size was reduced to $5 \mu\text{s}$. After ΔITD decreased to below $11 \mu\text{s}$, the step was further reduced to $2 \mu\text{s}$. Overall, 14 reversals were simulated for each analyzed frequency, and the ΔITDs were estimated by computing the average of the last 10 reversal points for each frequency.

4. Discrimination of ILD

The same ideal observer as in Sec. III D 3 was used to obtain the ILD discrimination threshold of the LSO model. The discrimination index is calculated only for a single band, the CF of which is nearest to the CF of the pure tone. The threshold detected by the observer will henceforth be denoted as ΔILD . Stimulus A had ILD equal to $\Delta\text{ILD}/2$ (lateralized to the right ear), while stimulus B had ILD equal to $-\Delta\text{ILD}/2$ (lateralized to the left ear). The simulation followed the 2AFC paradigm, as in the subjective experiment conducted by Yost and Dye (1988). The interval was considered as successfully predicted by the LSO model if d' was equal to or bigger than 0.95 and the model showed the correct direction of the lateral displacement. The selected d' criterion corresponds approximately to the 75% point on the psychometric function (Hacker and Ratcliff, 1979). The ΔILD value started at 1.5 dB in all cases and was decreased by steps of 0.25 dB until ΔILD of 0.4 dB was reached, after which the step size was reduced to 0.05 dB. Overall, 14 reversals were simulated for each analyzed frequency, and the ΔILDs were estimated by computing the average of the last 10 reversal points for each frequency.

5. Discrimination of phase warp

A procedure similar to the procedure used in the original study (Dietz *et al.*, 2008) was implemented to obtain the discrimination threshold between the phase warp and binaural uncorrelated noise. It was a two-down, one-up adaptive staircase procedure, which converges at the 70.7% correct point of the psychometric function (Levitt, 1971). The simulation followed the three-interval, three-alternative forced-choice paradigm. One of the intervals contained the phase-warp stimulus, while the other two intervals contained the binaurally uncorrelated NBNs of the same bandwidths as the phase warp.

In the first part of the experiment, the maximum beat frequency detectable by the MSO and LSO models was analyzed for phase warp with a bandwidth of either 550 Hz or 1100 Hz. The beat frequency started at 50 Hz and was increased/decreased by 15 Hz steps until the second reversal, after which a step size of 10 Hz was used for the next two

reversals before the final steps of 5 Hz was used for the remaining reversals. In the case of 1100 Hz bandwidth, the step size was doubled. Thirty reversals were simulated, and the mean was calculated from the last 20 values.

In the second part of the experiment, the smallest detectable modulation depth [see Eq. (12)] of phase-warp stimuli with fixed beat frequencies (10, 50, and 75 Hz) was analyzed. The modulation depth was first adjusted with a 4-dB step size until two reversals were observed, after which steps of 2 dB were used for the next two reversals, before the final step size of 1 dB was used for the remaining reversals.

The same ideal observer as in the article by [Dietz et al. \(2008\)](#) was used in this experiment. The outputs of the ITD and ILD central stages were transformed into the frequency domain. The magnitude spectra of the outputs, the f_c of which lie within the bandwidth of the stimuli, were averaged. The ideal observer compared the average spectra of the three stimuli and chose the stimulus with the most energy within the frequency bin corresponding to beat frequency f_b .

IV. RESULTS

In this section, the results of simulations of MSO and LSO models are compared with subjective data. In Figs. 6–10, the MSO data are depicted as filled diamonds, and the LSO data are depicted as filled triangles, both in light gray color connected by a solid line of the same color.

A. Lateralization of pure tones with IPD or ILD

The simulated data of the MSO model for pure tones with IPDs are depicted in Fig. 6. These data are compared with subjective data taken from [Yost \(1981\)](#). We only present a mode calculated by Yost from the responses of four subjects. The data shown here therefore does not indicate that some of the subjects reported ambiguous lateralization percepts for IPD larger than ± 90 deg. The MSO model data were multiplied by ten to match the subjective scale. Figure 6 is divided into five panels, each representing one pure-tone frequency (0.2, 0.5, 0.75, 1, and 1.5 kHz). The Pearson correlation coefficients between the simulated data and subjective data were calculated for each pure-tone frequency, and are shown in the corresponding panels. The coefficients indicate a high correlation between the simulated MSO model and the subjective data. The root-mean-square error RMSE = $\sqrt{(\sum_{j=1}^i (x_j - y_j)^2)/i}$ between the simulated data (x) and experimental data (y) was calculated for each pure-tone frequency, where i represents the total number of tested IPDs. In all cases, the RMSE is quite high due to deviations at extreme IPDs (-180 deg). This will be discussed in Sec. V (Discussion). The best agreement between the prediction and experimental data according to RMSE is at 0.2, 0.5, 0.75 and 1 kHz. In the case of 1 kHz, the performance of the model decays and shows a systematic decrease in sensitivity to IPDs toward higher frequencies. At 1.5 kHz, the model still follows the changes in IPD, but its output is almost damped.

The data for pure tones with ILDs are depicted in Fig. 7. The LSO model data were compared with subjective data taken from [Yost \(1981\)](#). The LSO model data were

multiplied by ten to match the subjective scale. Data for pure tones of frequencies 0.2, 0.5, 1, 2, and 5 kHz are displayed in five separate panels. As in the previous case, the Pearson correlation coefficients and RMSE between simulated and subjective data were calculated, and are shown in each panel. The coefficients indicate a high correlation between the simulated LSO model and the subjective data. The performance of the model is better than the human performance at 200 Hz. The best agreement between predictions and experimental data is at 0.5, 2, and 5 kHz. At 1 kHz, the model shows significantly lower sensitivity than the human subjects to ILD.

B. Lateralization of NBN with IPD or ILD

The mean subjective data and their standard deviations from the experiment with NBN with IPD or ILD are shown in Fig. 8. The subjective data for the NBNs with IPD are compared to the data from the MSO model in Figs. 8(A) and 8(B), and the subjective data for the NBNs with ILD are compared with the data from the LSO model in Figs. 8(C) and 8(D). The MSO and LSO model data were multiplied by ten to match the subjective scale. While the MSO data are qualitatively comparable with the subjective results, the LSO model deviates in the shape of the response as is discussed below. The MSO model goes back to zero lateralization for IPD ± 180 deg, which results from the fact that we change the IPD of each spectral component in the noise. For IPD of ± 180 deg, this manipulation creates two signals with an antiphase fine structure but with the same time domain envelope. Therefore, the listeners cannot use changes in the time domain envelope. In addition to pure tones with the opposite phase, these stimuli often create ambiguous lateralization on both sides. The RMSE and Pearson correlation coefficient are calculated for each NBN test case.

C. Discrimination of ITD

The discrimination data from the MSO model are shown in Fig. 9(A). The mean simulated data and their standard deviations are compared with the mean subjective data taken from [Brughera et al. \(2013\)](#). The simulated data shows good qualitative and quantitative agreement with the subjective data (S1 and S2) for higher frequencies. For lower frequencies, the ITD threshold of the model remains constant, while the subjective values increase. The additional point at 1460 Hz shows the highest frequency at which the MSO model is able to discriminate ITD in the pure tone.

D. Discrimination of ILD

A comparison between the LSO model simulations and subjective data is shown in Fig. 9(B). The mean subjective data were taken from [Yost and Dye \(1988\)](#). The model accounts for the loss of sensitivity at 1 kHz, but the discrimination of the pure tones for both lower and higher frequencies exceeded discrimination of the subjective data. The loss of sensitivity of the model at 1 kHz is due to a joint effect of the inhibitory delay, the first-order low-pass input filter, and the weighted moving average.

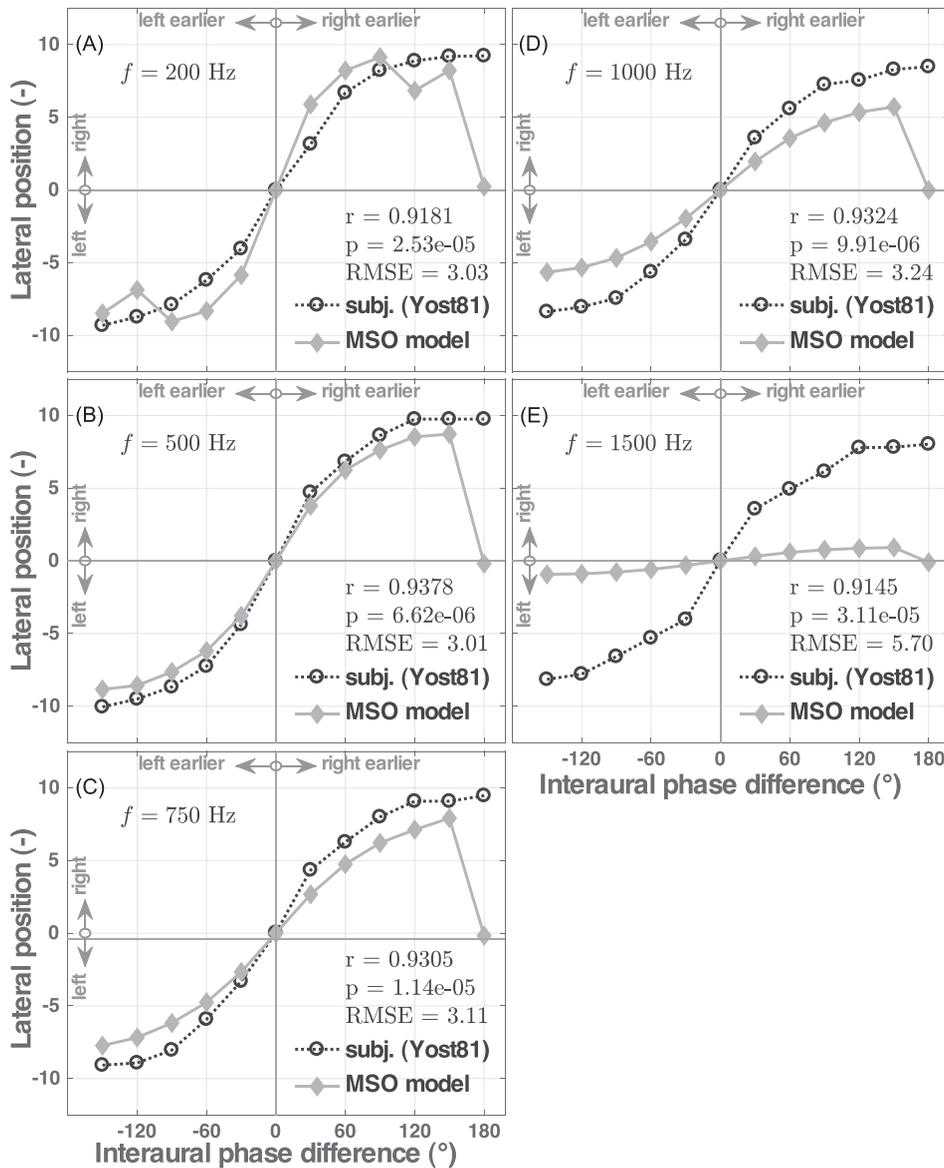


FIG. 6. Results of the lateralization experiment with pure tones with IPDs. The subjective data (Yost, 1981) are represented by circles (mode values), and are connected by a black dashed line. The responses of the MSO model are represented by diamonds connected by a solid gray line. (A)–(E) show the data for pure-tone frequencies of 0.2, 0.5, 0.75, 1, and 1.5 kHz, respectively.

E. Discrimination of phase warp

The predicted maximally detectable phase-warp beat frequencies are presented in Table I, along with psychophysical data collected by Dietz *et al.* (2008). The MSO model at the 500 Hz bandwidth exceeds the subjective results by about 50 Hz, but for the 1000 Hz bandwidth the model shows extreme sensitivity that is almost double the values for the subjective data. However, the LSO shows less sensitivity than the subjective data at 500 Hz bandwidth, but fit the subjective data for 1000 Hz bandwidth. Although the two models do not agree with the data quantitatively, they show qualitatively the same increasing trend of the discrimination threshold with increasing bandwidth as the subjective data. In addition, the temporal responses of both the MSO and LSO models to a phase warp with 8 Hz beat frequency are shown in Fig. 10(A). Both models showed apparent rotation of the sound image with the same frequency corresponding to the beat frequency of the phase warp. This behavior is in agreement with the subjective data of Dietz *et al.* (2008) and Siveke *et al.* (2008). The visible rotation diminished at about

10 Hz and changed to noise-like output, which is again in agreement with the subjective data (Dietz *et al.*, 2008; Siveke *et al.*, 2008).

The maximum detectable phase-warp modulation depth for the MSO and LSO models is depicted in Fig. 10(B). The subjective data were reproduced from the study by Dietz *et al.* (2008). As in the first part of the experiment, the MSO shows better discrimination than the subjects, and the slope of the curve also does not correspond with the subjective data. The LSO model shows a similar slope as the subjective data, but is about 2 dB less sensitive than the subjects at all of the three phase-warp frequencies.

V. DISCUSSION

The models of MSO and LSO showed a good match with human lateralization data for pure tones. The MSO model lateralization data deviated from the subjective data of Yost (1981) at IPDs between 120 and 180 deg, which may be due to the ambiguity of the laterality of such stimuli. This behavior was reported by subjects who heard two sound

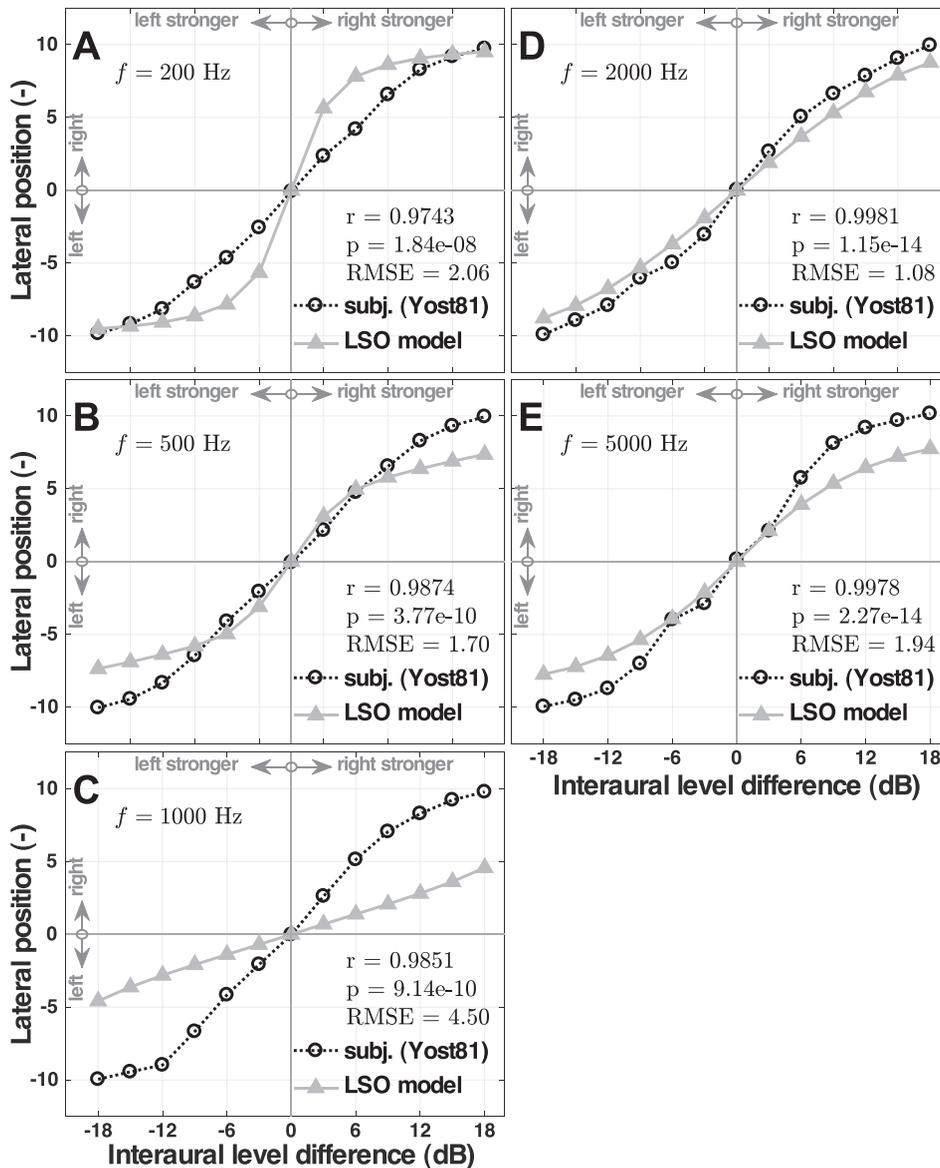


FIG. 7. Results of the lateralization experiment with pure tones with ILDs. The subjective data (Yost, 1981) are represented by circles (mode values), and are connected by a black dashed line. The responses of the LSO model are represented by triangles connected by a solid gray line. (A)–(E) show data for pure-tone frequencies of 0.2, 0.5, 1, 2, and 5 kHz, respectively.

images coming from opposite sides of the head (Yost, 1981). In their statistical analysis, these authors filtered correct output from the data using a mode instead of a mean value. A similar decrease in lateralization was reported by Sayers (1964), who used the mean of his subjective data. The MSO output at these extreme IPDs oscillates between maximum lateralization to the left and right sides, which ultimately decreases the mean output similarly to Sayers (1964), and may indicate directional ambiguity of this type of stimulus. Another discrepancy is the slope of the lateralization as a function of IPD for a pure tone of 1500 Hz. The change in the slope is mainly due to the rate-code principle used in the MSO model.

The performance of the model in the NBN stimuli around the center point (0 deg IPD or 0 dB ILD) in all tested setups is superior to the performance of the listeners. After this initial discrepancy, the slope change (second derivative) lags back and the performance of the model of more lateral localizations again matches the performance of the experimental subjects. As noise stimuli are more ecologically

relevant than pure tones, this poses an open question: Why is the performance of our model around the center point better than the performance of the experimental subjects? Strange behavior is evident in the case of the ILD experiment for low ILDs (range -3 to 3 dB), where the subjective lateralization remains around zero. This disagrees with Yost's pure-tone data (Yost, 1981). We investigated whether this stickiness to zero was caused by the way the ILD was induced to the stimuli. In an experiment, the ILD was added to one channel, while the other channel was left with the same amplitude. Five subjects participated. The results showed slightly higher stickiness to zero for both CFs of NBN. With two of the subjects, we also conducted several reduced experiments (fewer runs per test) in order to exclude any systematic error of the equipment or test methodology. Here is a full list of changes (each change was evaluated separately): the sound card was changed to RME Fireface 800 (Haimhausen, Germany); the headphones were changed to Sennheiser HD 280 Pro (Wedemark, Germany); the stimuli train was changed by removing the middle head reference,

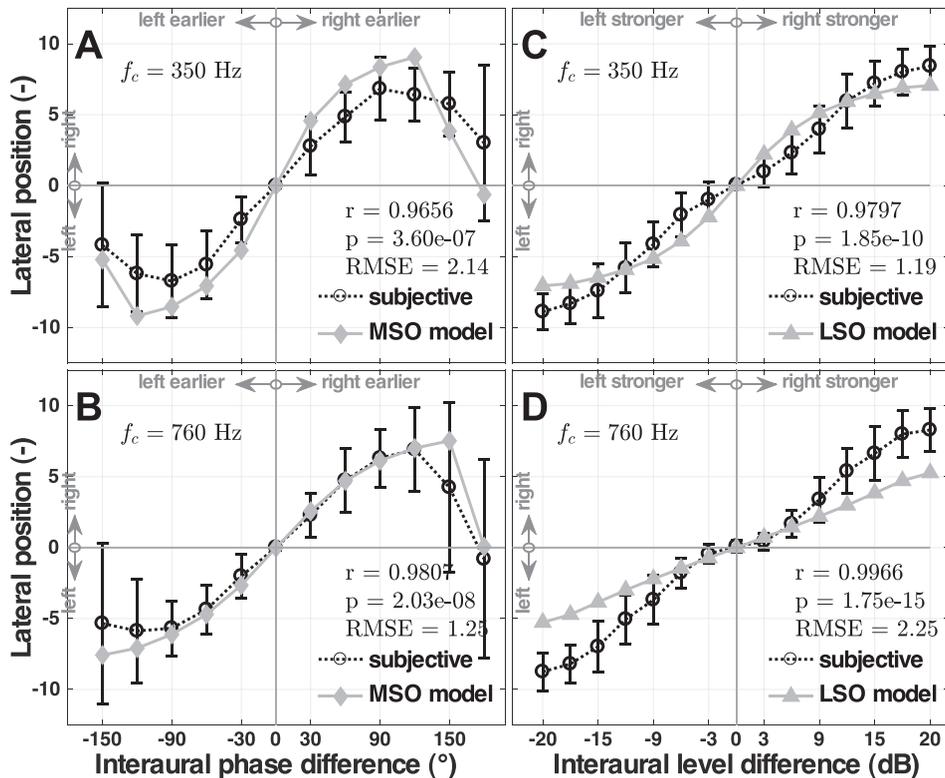


FIG. 8. Results of the lateralization experiment with NBN. The subjective data are represented by a circle (mean values) with a whisker (standard deviation) connected by a black dashed line. The responses of the MSO or LSO models are represented by a diamond or a triangle, respectively, connected by a solid gray line. (A) and (B) show the results of the NBN experiment with IPD, in comparison with the response of the MSO model, while (C) and (D) show the NBN experiment with ILD in comparison with the response of the LSO model. The top row shows the results for NBNs with 350 Hz f_c and the bottom row shows the results for NBNs with 760 Hz f_c .

increasing the stimuli length, and reducing the range of the ILD to ± 18 dB. In all cases, the overall shape of the lateralization curve remained the same. We hypothesized that the reason for the discrepancy might be in the envelope fluctuation of the NBN noise, which could deteriorate the ability of the hearing system to map small changes in the ILD to the lateral displacement. To test this hypothesis, we organized a quick informal test with 760 Hz pure tone. Six subjects participated, three of which had particular knowledge from earlier experiments, while the other three were naive. The ILD range was reduced to ± 6 dB. The subjective responses scale linearly with increasing ILDs, which follows the results of Yost (1981) and supports our hypothesis.

The ILD discrimination threshold of the LSO model shows a discrete reduction in sensitivity at 1 kHz, which is in line with the psychoacoustical data. In the model, this

phenomenon is caused by a joint effect of inhibitory delay, the first-order input filter of LSO, and the weighted moving average.

The ITD discrimination of the MSO fits the subjective data from Brughera *et al.* (2013) within their standard deviations. In contrast with the subjective data, however, the predicted data show no increase in the threshold at lower frequencies. The abrupt increase in the threshold with increasing frequency is mainly due to the LPF in the MSO model. With the peripheral filter only, there would be a shallower slope rise, and the model would not be able to decode the ITD at higher frequencies.

Both the MSO and LSO models show a trend with an increasing phase-warp discrimination threshold that is qualitatively similar to the subjective data. Quantitatively, the MSO model is more sensitive than the human subjects, and the LSO

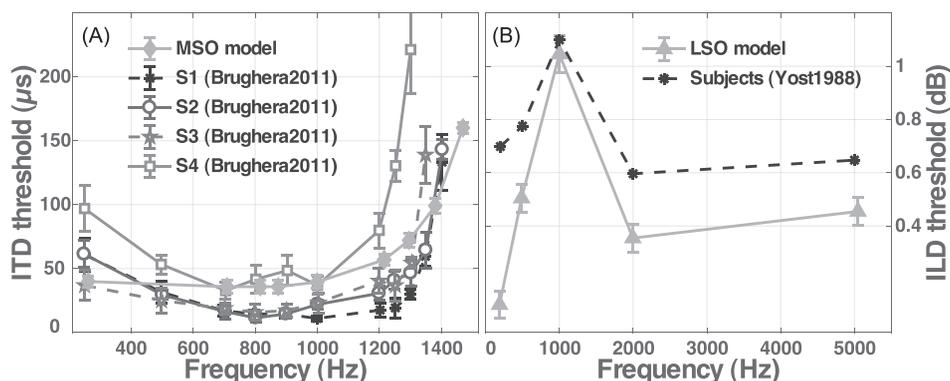


FIG. 9. (A) The results of the ITD discrimination experiment; the MSO model data are represented by a gray diamond with a whisker, which represents the standard deviations, and are interconnected by a solid gray line. The mean subjective data from four subjects from Brughera *et al.* (2013) are depicted as circles, hexagons, stars, and squares. (B) The results of the ILD discrimination experiment, where the LSO model data are represented by gray triangles and the subjective data reproduced from Yost and Dye (1988) are represented as black dots.

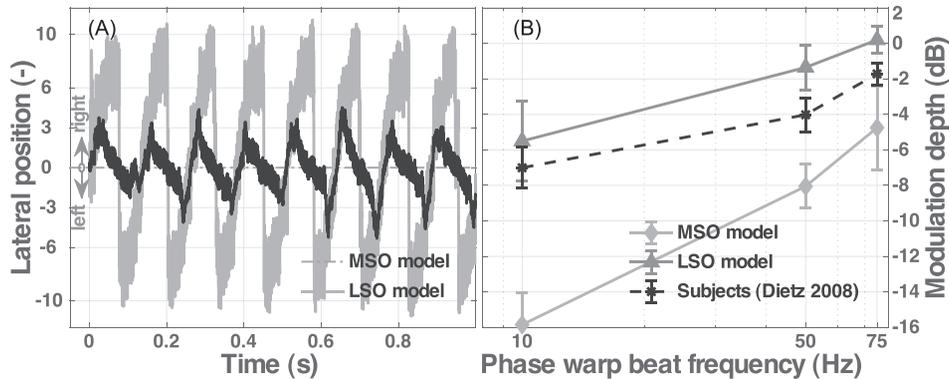


FIG. 10. (A) Transient responses of the MSO and LSO models to a phase-warp stimulus with the beat frequency $f_b = 8$ Hz. (B) The results of the experiment with a phase warp with variable modulation depth. The response of the MSO model is represented by diamonds connected by a solid gray line, and the response of the LSO model is represented by triangles connected by a gray dotted line. The subjective data reproduced from Dietz *et al.* (2008) are shown by a black dashed line. The standard deviations are depicted as bars above each datapoint.

model is less sensitive than human subjects in the case of 500 Hz bandwidth, but shows good fit for 1000 Hz bandwidth.

In the modulated phase-warp discrimination task, the MSO model again shows better performance than is shown by the subjective data. Surprisingly, the LSO model matches the subjective data qualitatively, but is about 2 dB less sensitive. This is in line with the neurophysiological data about LSO neurons being sensitive to envelope-ITDs in amplitude-modulated signals (Joris and Yin, 1995). These two results indicate a possible theoretical contribution of MSO and LSO to the decoding of phase-warp stimuli.

The models of MSO and LSO presented here are the rate-code based models like, for example, the models of Dietz *et al.* (2008), Pulkki and Hirvonen (2009), Takanen *et al.* (2014), van Bergeijk (1962), and von Békésy (1930). The MSO model is most comparable with the models of Pulkki and Hirvonen (2009) and Takanen *et al.* (2014). In line with physiology, these MSO models account for the shorter arrival time of the inhibitory signal from the contralateral ear. This is not incorporated in the models of Dietz *et al.* (2008), van Bergeijk (1962), and von Békésy (1930). The presented LSO model assumes two inputs with slightly different time delays (shorter for the ipsilateral ear). These inputs are then subtracted. Generally this approach is similar to the LSO model of Dietz *et al.* (2011), Pulkki and Hirvonen (2009), and Takanen *et al.* (2014). The only difference is that the model of Dietz *et al.* (2011) does not incorporate the different delays of signals coming from the ipsilateral and contralateral ears.

Our models of MSO and LSO demonstrate that with a relatively simple neurophysiologically inspired signal processing design, it is possible to obtain a performance comparable to human listeners in lateralization tasks. In addition, the ITD and ILD central stages' outputs give values directly representing the subjective lateralization, which is advantageous to the previous rate-code models (Pulkki and

Hirvonen, 2009; Takanen *et al.*, 2014). The model further supports the hypothesis that the information about ITD and ILD is coded with a rate-code instead of a place code. However, it should be noted that place-code models based on the Jeffress delay line can account for all of the phenomena shown in this study, and even for more complex phenomena (Braasch *et al.*, 2013; Breebaart *et al.*, 2001; Colburn, 1977; Lindemann, 1986; Prokopiou *et al.*, 2017).

The information about lateralization is processed independently for the LSO and MSO models. However, it is unlikely that the processing in the inferior colliculus would be different for MSO and LSO. The overall structures of the central stages tend to follow the functional aspect rather than the physiology. This also limits the possibility to carry out reliable experimental cue-trading tests in which ILD is compensated by ITD and vice versa. The performance of our LSO model is affected if the arrival time between the signals in the left and right ears is changed, as the different time delays between the neural signal in the contralateral and ipsilateral ears in the model affect the discrimination threshold for ILD. The effect of ILD on the outcome of the MSO model is minimal. Moreover, in the presented implementation of both models one has to know beforehand the stimulus spectral properties and which binaural cue it contains to obtain proper results. If these parameters were unknown, one possible solution was presented in a pilot study by Koshkina and Bouse (2017), which has shown that if the former versions of presented models are combined with the K -nearest neighbour (KNN) or ANN learning algorithms, they can be used for localization tasks with a head-related transfer function (HRTF), which contains both ITD and ILD information. Mean values in one time frame (variable length) of the output signals from the MSO and LSO models were processed separately by machine learning algorithms. Signals in each critical band were processed separately, and in the case of the MSO model only the critical bands below 1.5 kHz were used. The KNN and ANN were trained on speech samples, which were filtered by HRTFs for different azimuths between -90 and 90 deg. The algorithm was then tested by using another set of speech samples created by the same HRTFs. The decision device was based on a simple averaging of the predicted azimuths from the MSO and LSO parts, and if the deviation between the MSO and LSO predictions was more than 20 deg only the prediction from the LSO was taken into account. The emphasis on the LSO predictions aroused from the simulations where the prediction from the LSO was more

TABLE I. The results of the first part of the phase-warp experiment. The table shows mean subjective, MSO, and LSO data with standard deviations for two different phase-warp bandwidths.

Phase-warp bandwidth	Mean subjective data (Dietz <i>et al.</i> , 2008)	MSO model	LSO model
550 Hz	96 ± 15 Hz	143.5 ± 15 Hz	61 ± 17.2 Hz
1100 Hz	219 ± 30 Hz	492 ± 22 Hz	211 ± 48 Hz

stable than the one from the MSO. The algorithm(s) perfectly localized the azimuths near 0 deg. The accuracy decreased with increasing absolute value of the azimuth.

VI. CONCLUDING REMARKS

Two-channel binaural models of MSOs and LSOs were designed according to three criteria. The first criterion was that the models should take into account current neurophysiological findings (Brand *et al.*, 2002; Grothe, 1994, 2003; Joris, 1996; Joris and Yin, 1995; Roberts *et al.*, 2013; Tollin and Yin, 2005). Therefore, both models are based on the rate-code principle. In addition, the MSO model is most sensitive if a pure tone in the contralateral ear is delayed by 50-deg IPD [Fig. 3(A)], which agrees with neurophysiological data of Grothe (2003). And with delayed broadband noise, the MSO output shows the broadest peak for low CFs at high ITDs, contrary for high CFs it shows a sharper peak at low ITDs [see Fig. 3(B)], which agrees with the neurophysiological data of McAlpine *et al.* (2001).

The second criterion was that the models should give a quantitative representation of the subjective lateralization. If complemented with phenomenological central stages of ITD and ILD, the models' predictions show a good match with subjective data from literature: pure-tone lateralization (Yost, 1981), ITD and ILD discrimination (Brughera *et al.*, 2013; Yost and Dye, 1988), and phase-warp discrimination (Dietz *et al.*, 2008). In addition, the models also predict NBN lateralization data obtained by the authors. In the case of ILD discrimination, the LSO model predicts an experimentally observed (Yost and Dye, 1988) decrease in sensitivity around 1 kHz.

The third criterion was that everything mentioned in criteria 1 and 2 should be achieved with low structural and computational complexity. From the computational point of view, it is hard to make an evaluation without having all competing models in the same test scenarios. However, the computational performance of both models can be further increased by reducing the peripheral filters spacing to 1 ERB in a cost of reduced performance in the phase-warp detection task. Similarly, structural complexity is a highly subjective measure, so we have left the reader to decide for himself/herself whether this criterion has been met.

The presented rate-code models provide a simpler approach than their predecessors Dietz *et al.* (2008), Encke and Hemmert (2018), Pulkki and Hirvonen (2009), and Takanen *et al.* (2014), and are still able to follow the fundamentals of rate-code models and accurately predict ITD and ILD based lateralization results. On the other hand, we have to admit that the model in the paper was tested by using simple stimuli, whereas the aforementioned models were tested in more challenging listening situations. In addition, the models from Jeffress' family are able to predict subjective data even in more complex listening scenarios, for example, with presence of sound distractors and room echoes (Faller and Merimaa, 2004).

Overall, both models presented here serve as a possible piece in the puzzle surrounding the processing of binaural hearing in the mammalian brain. And, due to their relatively

low complexity and good performance, one can use the presented models in real-time applications.

ACKNOWLEDGMENTS

This research was supported by the Grant Agency of the Czech Technical University in Prague, Grant No. SGS17/190/OHK3/3 T/13, and the European Commission Danube Project, Grant No. DS-2016-0026, Plasticity of (Spatial) Processing in Normal and (C)ochlear (I)mplant Hearing (SpaCI). We would like to thank Z. Bures for helpful suggestions, three anonymous reviewers for their valuable comments, and all the subjects who voluntarily participated in the experiments. Special thanks to T. Reks and S. Uperjaarda.

¹<http://mmtg.fel.cvut.cz/rate-code-model/> (Last viewed December 17, 2018).

- ANSI (1996). S3.6-1996, *American National Specifications for Audiometers* (American National Standard Institute, New York).
- Bernstein, L. R., and Trahiotis, C. (1985). "Lateralization of low-frequency complex wave-forms: The use of envelope-based temporal disparities," *J. Acoust. Soc. Am.* **77**, 1868–1880.
- Blauert, J. (1997). *Spatial Hearing—The Psychophysics of Human Sound Localization* (MIT Press, Cambridge, MA).
- Bouse, J., and Vencovsky, V. (2015). "Two-channel models of medial and superior olive based on psychoacoustics," *BMC Neuroscience* **16**(Suppl 1), P276.
- Braasch, J., Clapp, S., Parks, A., Pastore, T., and Xiang, N. (2013). "A binaural model that analyses acoustic spaces and stereophonic reproduction systems by utilizing head rotations," in *The Technology of Binaural Listening*, edited by J. Blauert (Springer, Berlin), pp. 201–223.
- Brand, A., Behrend, O., Marquardt, T., McAlpine, D., and Grothe, B. (2002). "Precise inhibition is essential for microsecond interaural time difference coding," *Nature* **417**, 543–547.
- Breebaart, J., van de Par, S., and Kohlrausch, A. (2001). "Binaural processing model based on contralateral inhibition. I. Model structure," *J. Acoust. Soc. Am.* **110**, 1074–1088.
- Brughera, A., Dunai, L., and Hartmann, W. M. (2013). "Human interaural time difference thresholds for sine tones: The high-frequency limit," *J. Acoust. Soc. Am.* **133**, 2839–2855.
- Bures, Z. (2012). "The stochastic properties of input spike trains control neuronal arithmetic," *Biol. Cybern.* **106**, 111–122.
- Bures, Z., and Marsalek, P. (2013). "On the precision of neural computation with interaural level differences in the lateral superior olive," *Brain Res.* **1536**, 16–26.
- Cherry, E. C., and Sayers, B. M. (1956). "Human cross-correlator—A technique for measuring certain parameters of speech perception," *J. Acoust. Soc. Am.* **28**, 889–895.
- Colburn, H. S. (1977). "Theory of binaural interaction based on auditory-nerve data. II. Detection of tones in noise," *J. Acoust. Soc. Am.* **61**, 525–533.
- Colburn, H. S. (1978). "Models of binaural interaction," in *Handbook of Perception, Vol. IV*, edited by E. Carterette and M. Friedman (Academic, San Diego, CA), pp. 467–518.
- Dietz, M., Ewert, S. D., and Hohmann, V. (2009). "Lateralization of stimuli with independent fine-structure and envelope-based temporal disparities," *J. Acoust. Soc. Am.* **125**, 1622–1635.
- Dietz, M., Ewert, S. D., and Hohmann, V. (2011). "Auditory model based direction estimation of concurrent speakers from binaural signals," *Speech Commun.* **53**, 592–605.
- Dietz, M., Ewert, S. D., Hohmann, V., and Kollmeier, B. (2008). "Coding of temporally fluctuating interaural timing disparities in a binaural processing model based on phase differences," *Brain Res.* **1220**, 234–245.
- Encke, J., and Hemmert, W. (2018). "Extraction of inter-aural time differences using a spiking neuron network model of the medial superior olive," *Front. Neurosci.* **12**, 140.
- 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**, 3075–3089.

- Gaik, W. (1993). "Combined evaluation of interaural time and intensity differences: Psychoacoustic results and computer modeling," *J. Acoust. Soc. Am.* **94**, 98–110.
- Goode, R. L., Killion, M., Nakamura, K., and Nishihara, S. (1994). "New knowledge about the function of the human middle ear: Development of an improved analog model," *Am. J. Otol.* **15**, 145–154.
- Grantham, D. W. (1984). "Interaural intensity discrimination: Insensitivity at 1000 Hz," *J. Acoust. Soc. Am.* **75**, 1191–1194.
- Grothe, B. (1994). "Interaction of excitation and inhibition in processing of pure tone and amplitude-modulated stimuli in the medial superior olive of the mustached bat," *J. Neurophysiol.* **71**, 706–721.
- Grothe, B. (2003). "New roles for synaptic inhibition in sound localization," *Nat. Rev. Neurosci.* **4**, 540–550.
- Grothe, B., Pecka, M., and McAlpine, D. (2010). "Mechanisms of sound localization in mammals," *Physiol. Rev.* **90**, 983–1012.
- Hacker, M. J., and Ratcliff, R. (1979). "A revised table of d' for M -alternative forced choice," *Percept. Psychophys.* **26**, 168–170.
- Hartmann, W. M., Rakerd, B., and Crawford, Z. D. (2016). "Transaural experiments and a revised duplex theory for the localization of low-frequency tones," *J. Acoust. Soc. Am.* **139**, 968–985.
- Jeffress, L. A. (1948). "A place theory of sound localization," *J. Comp. Physiol. Psychol.* **41**, 35–39.
- Joris, P. X. (1996). "Envelope coding in the lateral superior olive. II. Characteristic delays and comparison with responses in the medial superior olive," *J. Neurophysiol.* **76**, 2137–2156.
- Joris, P. X., and Yin, T. C. T. (1995). "Envelope coding in the lateral superior olive. I. Sensitivity to interaural time differences," *J. Neurophysiol.* **73**, 1043–1062.
- Klump, R. G., and Eady, H. R. (1956). "Some measurements of interaural time difference thresholds," *J. Acoust. Soc. Am.* **28**, 859–860.
- Koshkina, E., and Bouse, J. (2017). "Localization in static and dynamic hearing scenarios: Utilization of machine learning and binaural auditory model," in *Proc. of 21th International Scientific Student Conference POSTER 2017*, Czech Technical University, Prague, pp. 1–5.
- Levitt, H. (1971). "Transformed up-down methods in psychoacoustics," *J. Acoust. Soc. Am.* **49**, 467–477.
- Lindemann, W. (1986). "Extension of a binaural cross-correlation model by contralateral inhibition. II. The law of the first wave front," *J. Acoust. Soc. Am.* **80**, 1623–1630.
- Lopez-Poveda, E. A., and Meddis, R. (2001). "A human nonlinear cochlear filterbank," *J. Acoust. Soc. Am.* **110**, 3107–3118.
- Marquardt, T., and McAlpine, D. (2007). "A π -limit for coding ITDs: Implications for binaural models," in *Hearing—From Sensory Processing to Perception*, edited by B. Kollmeier (Springer, Berlin), pp. 407–416.
- McAlpine, D., and Grothe, B. (2003). "Sound localization and delay lines—Do mammals fit the model?," *Trends. Neurosci.* **26**, 347–350.
- McAlpine, D., Jiang, D., and Palmer, A. R. (2001). "A neural code for low-frequency sound localization in mammals," *Nat. Neurosci.* **4**, 396–401.
- Mills, A. W. (1960). "Lateralization of high frequency tones," *J. Acoust. Soc. Am.* **32**, 132–134.
- Moore, B. C. J. (2003). *An Introduction to the Psychology of Hearing*, 5th ed. (Academia, San Diego).
- Moore, B. C. J., and Glasberg, B. R. (1983). "Suggested formulae for calculating auditory-filter bandwidths and excitation patterns," *J. Acoust. Soc. Am.* **74**, 750–753.
- Pralong, D., and Carlile, S. (1996). "The role of individualized headphone calibration for the generation of high fidelity virtual auditory space," *J. Acoust. Soc. Am.* **100**, 3785–3793.
- Prokopiou, A., Moncada-Torres, A., Wouters, J., and Francart, T. (2017). "Functional modelling of interaural time difference discrimination in acoustical and electrical hearing," *J. Neural. Eng.* **14**, 1–21.
- Pulkki, V., and Hirvonen, T. (2009). "Functional count-comparison model for binaural decoding," *Acta Acust. Acust.* **95**, 883–900.
- Rayleigh, O. M. (1907). "On our perception of sound direction," *Philos. Mag.* **13**, 214–232.
- Roberts, M. T., Seeman, S. C., and Golding, N. L. (2013). "A mechanistic understanding of the role of feedforward inhibition in the mammalian sound localization circuitry," *Neuron* **78**, 923–935.
- Sakitt, B. (1973). "Indices of discriminability," *Nature* **241**, 133–134.
- Salminen, N. H., Tiitinen, H., Yrttiaho, S., and May, P. J. C. (2010). "The neural code for interaural time difference in human auditory cortex," *J. Acoust. Soc. Am.* **127**, EL60–EL65.
- Sayers, B. M. (1964). "Acoustic-image lateralization judgments with binaural tones," *J. Acoust. Soc. Am.* **36**, 923–926.
- Siveke, I., Ewert, S. D., Grothe, B., and Wiegand, L. (2008). "Psychophysical and physiological evidence for fast binaural processing," *J. Neurosci.* **28**, 2043–2052.
- Søndergaard, P., and Majdak, P. (2013). "The auditory modeling toolbox," in *The Technology of Binaural Listening*, edited by J. Blauert (Springer, Berlin), pp. 33–56.
- Stern, R. M., and Colburn, H. S. (1978). "Theory of binaural interaction based in auditory-nerve data. IV. A model for subjective lateral position," *J. Acoust. Soc. Am.* **64**, 127–140.
- Stevens, S. S., and Newman, E. B. (1936). "The location of actual sources of sound," *Am. J. Psych.* **48**, 297–306.
- Takanen, M., Santala, O., and Pulkki, V. (2014). "Visualization of functional count-comparison-based binaural auditory model output," *Hear. Res.* **309**, 147–163.
- Tollin, D. J. (2003). "The lateral superior olive: A functional role in sound source localization," *Neuroscientist* **9**, 127–143.
- Tollin, D. J., and Yin, T. T. (2005). "Interaural phase and level difference sensitivity in low-frequency neurons in the lateral superior olive," *J. Neurosci.* **25**, 10648–10657.
- van Bergeijk, W. A. (1962). "Variation on a theme of Bekesy: A model of binaural interaction," *J. Acoust. Soc. Am.* **34**, 1431–1437.
- Vítek, S., Klíma, M., Husník, L., and Špírk, D. (2011). "New possibilities for blind people navigation," in *International Conference on Applied Electronics*, Pilsen, pp. 405–408.
- von Békésy, G. (1930). "Zur Theorie des Hörens. Über das Richtungshören bei einer Zeitdifferenz oder Lautstärkenungleichheit der beiderseitigen Schalleinwirkungen" ("On the theory of hearing: On directional hearing with a time difference or inequality of loudness as a sound effect between the two sides"), *Physik Z.* **31**, 824–835.
- Weiss, T. F., and Rose, C. (1988). "A comparison of synchronization filters in different auditory receptor organs," *Hear. Res.* **33**, 175–179.
- Wightman, F. L., and Kistler, D. J. (1992). "The dominant role of low-frequency interaural time differences in sound localization," *J. Acoust. Soc. Am.* **91**, 1648–1661.
- Yost, W. A. (1981). "Lateral position of sinusoids presented with interaural intensive and temporal differences," *J. Acoust. Soc. Am.* **70**, 337–409.
- Yost, W. A., and Dye, R. H., Jr. (1988). "Discrimination of interaural differences of level as a function of frequency," *J. Acoust. Soc. Am.* **83**, 1846–1851.
- Zwislocki, J., and Feldman, R. S. (1956). "Just noticeable differences in dichotic phase," *J. Acoust. Soc. Am.* **28**, 860–864.