

A framework for testing and comparing binaural models

Mathias Dietz^{a,1}, Jean-Hugues Lestang^b, Piotr Majdak^c, Richard M Stern^d, Torsten Marquardt^e, Stephan D Ewert^f, William M Hartmann^g, Dan F M Goodman^b

a) National Centre for Audiology, Western University, London, ON, Canada. b) Department of Electrical and Electronic Engineering, Imperial College London, London, United Kingdom. c) Institut für Schallforschung, Österreichische Akademie der Wissenschaften, Wien, Austria. d) Carnegie Mellon University, Pittsburgh, PA, United States. e) Ear Institute, University College London, London, United Kingdom. f) Medizinische Physik, Universität Oldenburg, Oldenburg, Germany. g) Michigan State University, East Lansing, MI, United States.

1) Corresponding author email: mdietz@uwo.ca

Abstract

Auditory research has a rich history of combining experimental evidence with computational simulations of auditory processing in order to deepen our theoretical understanding of how sound is processed in the ears and in the brain. Despite significant progress in the amount of detail and breadth covered by auditory models, for many components of the auditory pathway there are still different model approaches that are often not equivalent but rather in conflict with each other. Similarly, some experimental studies yield conflicting results which has led to controversies. This can be best resolved by a systematic comparison of multiple experimental data sets and model approaches. Binaural processing is a prominent example of how the development of quantitative theories can advance our understanding of the phenomena, but there remain several unresolved questions for which competing model approaches exist. This article discusses a number of current unresolved or disputed issues in binaural modeling, as well as some of the significant challenges in comparing binaural models with each other and with the experimental data. We introduce an auditory model framework, which we believe can become a useful infrastructure for resolving some of the current controversies. It operates models over the same paradigms that are used experimentally. The core of the proposed framework is an interface that connects three components irrespective of their underlying programming language: The experiment software, an auditory pathway model, and task-dependent decision stages called artificial observers that provide the same output format as the test subject.

Keywords:

Auditory models; Reproducible research; Binaural processing; Psychoacoustics; Interaural Time Differences

1. Introduction

Binaural hearing is vital for localizing sound sources (Rayleigh, 1907) as well as for selectively focusing on a target sound while suppressing sound from other directions (Cherry, 1953). Interaction in the brainstem of auditory information from the left and the right ears, which forms the basis of binaural hearing, is one of the fundamental components and has a long history of model-inspired research (e.g., Jeffress, 1948; von Békésy, 1930). Essentially, the auditory system extracts binaural information including the interaural time difference (ITD) and the interaural level difference (ILD) between the sound reaching the left and the right ear. Especially for ITD extraction, highly accurate neural circuits have evolved (e.g., Stange-Marten et al., 2017). A functional understanding of binaural circuits is important for explaining the very rich sets of behavioral and neural data, as well as for the development and fitting of binaural assistive hearing devices (e.g., Dietz, 2016). Furthermore, a biophysical understanding is becoming increasingly relevant not only for basic neuroscience, but also for understanding the limitations of binaural hearing with cochlear implants (e.g., Chung et al., 2015) and, in the probable future, for the diagnosis and treatment of hearing disorders of neuronal origin.

Given this long history of binaural modelling, it is natural to ask what is today's state-of-the-art in models of binaural processing. What can be explained, and what are the remaining phenomena that are not easily understood? The dominant modeling perspective has been a framework that was proposed by Jeffress (1948): Coincidence-detecting neurons, aligned along (axonal) delay lines. The delay elements compensate for the external ITD somewhere along the delay line and coincidence is detected. This was generally thought to be in line with physiology as evidenced by the existence of axonal delays and an array of neural coincidence detectors along these axons in the barn owl (Carr and Konishi, 1990). This array of coincidence detecting neurons can be considered to be a physiologically-based representation of the interaural cross-correlation of the signals after (monaural) processing by the peripheral auditory system. The Jeffress coincidence-counting hypothesis had been generally accepted and further quantified (e.g., Colburn, 1977; Sayers and Cherry, 1957) and has made its way into many general neuroscience or audiology textbooks. To date, cross-correlation-based models with the putative interaural delay line are arguably the best predictors of human psychoacoustical performance covering a broad range of lateralization data (Bernstein and Trahiotis, 2012; Lindemann, 1986; Shackleton et al., 1992; Stern and Colburn, 1978; Stern and Shear, 1996; Trahiotis and Stern, 1989) and dichotic tone-in-noise detection results (e.g., Bernstein and Trahiotis, 2017; Breebaart et al., 2001a, 2001b, 2001c; Colburn, 1977).

While cross-correlation-based models have generally been successful in predicting psychoacoustical data, the results of some more recent physiological studies using gerbils and guinea pigs (e.g., Brand et al., 2002; McAlpine et al., 2001; Pecka et al., 2008) do not support the existence of a similar network in those species and cast doubt on its existence in humans. Despite some success of new models which lack the array of different axonal delays (Dietz et al., 2009; Marquardt and McAlpine, 2009; Takanen et al., 2014), quantitative accuracy is often better with models that include delay lines (see, e.g., Marquardt and McAlpine, 2009 for a comparative study). It remains unclear if the newer models based on more recent physiological results have been less successful in predicting the results of a wider range of perceptual data because they represent a fundamentally incorrect conceptual formulation, or simply because the newer models have not yet experienced the refinement over multiple decades that the delay-line-based models have enjoyed.

The scope of this article is twofold. First, we review and discuss binaural modelling with respect to its potentials and limitations for comparative studies. Second, we introduce a modelling framework that specifically aims at optimizing comparative auditory modelling studies. Before the review-oriented Secs. 2 to 4, two central components of the proposed framework are worth mentioning upfront. The first

component is the concept of the artificial observer. This is a general concept which only requires that the decision stage behind a model provides the same output format as a real observer does. In a psychoacoustical n -alternative task this is typically the selection of one of possibly multiple response alternatives. In physiology this may be a spike train. We believe that this approach offers the best comparison of model predictions to experimental results. This approach has been highly successful for modeling monaural perception (e.g., Dau et al., 1997a,b; Jepsen et al., 2008) but has only rarely been employed in binaural modeling (e.g., Breebaart *et al.*, 2001a,b,c). Models with an artificial observer can be tested in the exact same way as a human subject, typically with the same measurement software routines. Fig. 1 illustrates an experimental program operating in the same way with real subjects and with one or more models. In addition, responses by an artificial observer can be analyzed and displayed in the exact same way as the experimental data. This is particularly useful in adaptive experimental procedures, where the results may depend on the experimental paradigm. Therefore, some measurement software routines for psychoacoustics already enable switching between a real subject and an artificial observer (e.g., Ewert, 2013). The concept of the artificial observer is related to the concept of the ideal decision maker (e.g. Colburn, 1977). The ideal decision maker combines information from all cues available in a stimulus. This represents an artificial observer in a single-interval task. In multi-interval forced-choice tasks, however, the artificial observer is typically comparing across intervals.

The second central component of the framework is the experiment-to-model interface. It is often difficult to connect an experiment with a model, let alone to connect several experiments with several models in multiple combinations. In Sec. 5.4 we will describe an interface protocol and provide code (see <http://github.com/model-initiative>) that performs this connection in a platform- and programming-language-independent way.

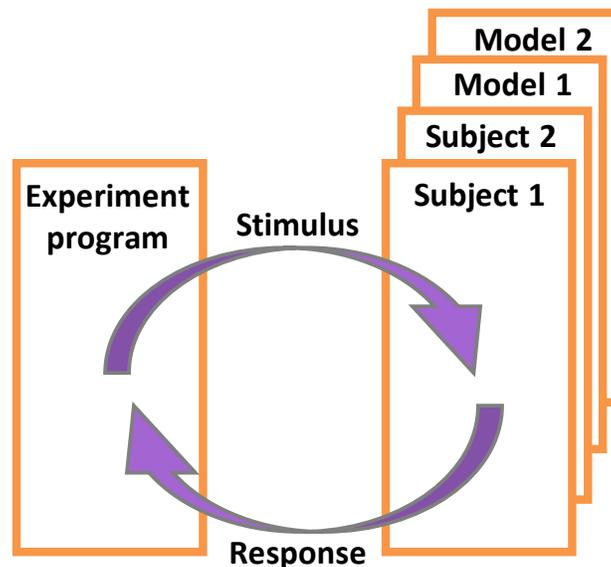


Figure 1. Concept of a model with “artificial observer”. The Experiment program is operating in the same way with one or several models as it does with the real subjects. The multiple “observer” blocks on the right indicate that the observer can be changed, while the experimental program remains unchanged.

2. Selected open questions of binaural processing

Binaural modelling is particularly suited for comparative studies, because there are several aspects of binaural processing for which multiple competing hypotheses are currently being debated. This section reviews five exemplary aspects of binaural processing and lists selected models representing competing concepts. Where possible, suggestions for potential comparative model studies are made.

2.1 Why neurons respond maximally at non-zero ITDs

The distribution of ITDs at which neurons have their maximum response is referred to as the $p(\tau)$ function (Stern and Colburn, 1978). There is some debate over the question of whether the highest density is near $\tau=0$, as assumed in many historically important models (e.g., Stern and Colburn 1978), or at a non-zero delay corresponding approximately to an interaural phase difference (IPD) of 45 deg (McAlpine et al., 2001). However, all distributions contain neurons that respond maximally at non-zero ITDs. The respective ITD of maximum response is typically referred to as best delay (e.g., Joris and Yin, 2007).

Open questions include the distribution of best delays in humans (e.g., Harper et al., 2014; Thompson et al., 2006; van der Heijden and Trahiotis, 1999) and the cause of the non-zero best delay (see, e.g., Joris and Yin, 2007, for a comparative review). It is beyond the scope of this review to explain the various mechanisms proposed for the internal delays such as interaural mismatches in CF (Schröder, 1977; Shamma et al., 1989), axonal delay lines (Carr and Konishi, 1990), synaptic inhibition (Batra et al., 1997; Brand et al., 2002), or an interaction between synaptic inputs and the intrinsic neural conductances (Zhou et al., 2005). Given the complexity of this topic, a comparative model study with such a focus may be a promising approach to clarify this still-open question. In many of these cases it has been challenging to identify a set of experimental data that can be used to differentiate between competing proposed mechanisms.

2.2 Decoding neural responses for sound localization

A related but not identical issue under dispute is how the perceived location of a sound source is encoded in the brain. Most of the physiological binaural literature is centered on the processing concepts of primary binaural neurons in the brainstem similar to our Sec. 2.1. However, in the current subsection the question is how higher stages process the information from the primary layer. Before describing several of the conflicting model concepts, it is important to understand that in many cases modelers suggest the proposed mechanisms either without claiming or very weakly claiming that their model explicitly describes how the brain really processes the information. To a much greater extent than in conjunction with the last topic in Sec. 2.1 the different approaches below should be understood to be highly simplified conceptualizations of much more complex and variable brain mechanisms. Their primary goal and merit is to offer a functional understanding.

Most delay-line-based models postulate a place code in some form, in that the neuron's best delay compensates for the external ITD. It is then assumed that higher pathway stages have learned to assign a specific sound direction or ITD to each neuron and identify which neuron has the highest response rate (Jeffress, 1948; Shackleton et al., 1992). The two key aspects of place coding are (1): The low and moderate response rates of all other neurons are ignored and (2): the function of the best delay is ITD compensation. This concept is sketched in Fig. 2A.

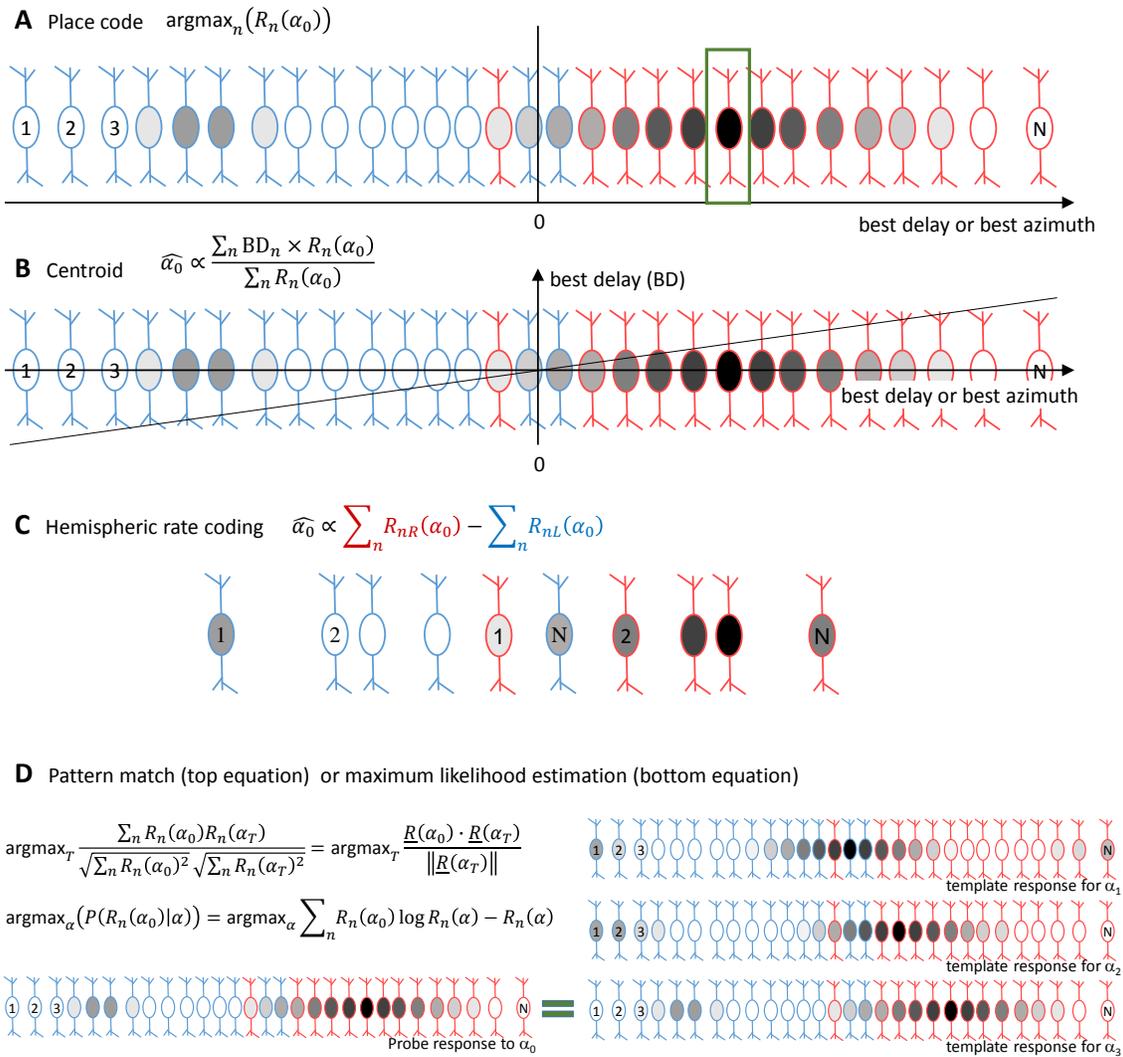


Figure 2. Different approaches to decoding the azimuthal sound direction or sound image lateralization from the response of binaural neurons. A probe stimulus is presented with direction α_0 . The estimated direction is $\hat{\alpha}_0$ and the response rate of the n -th neuron R_n for direction α_0 is thus $R_n(\alpha_0)$. For the purpose of this illustration, direction α and ITD can be seen interchangeably. The cell body color illustrates the response rate R_n : White = no response; black = maximum response. Axon color indicates brain hemisphere in which the neuron resides: Blue = left; Red = right hemisphere. A: Place code (Jeffress, 1948). Direction is encoded by the label of the most responsive neuron(s). B: Centroid (Blauert and Cobben, 1978; Stern and Colburn, 1978). Direction or lateralization is proportional to the best-delay-weighted neural response across neurons, computing in effect the center of mass of the response as a function of best delay. C: Hemispheric rate code (Bekesy 1930): Direction or lateralization is proportional to the rate difference or ratio between all binaural neurons from each hemispheres. Tuning of neurons is ignored. A smaller number of neurons is typically used, but the number and distribution of neurons is not in the focus of Sec 2.2. D: Non-parametric population codes such as pattern matching (Goodman et al. 2013) or maximum likelihood estimation (Day and Delgutte, 2013). The brain has learned the response of all neurons as a function of direction. For a probe stimulus it calculates the direction where the response of all individual neurons is in best agreement with the learned response.

A second historic approach to decoding azimuthal direction or sound image lateralization is based on the hemispheric rate difference of all neurons (van Bergeijk, 1962; von Békésy, 1930). For a rate code, typically

fewer neurons are required and the function of their non-zero best delay is to generate a slope in the hemispheric response rate difference that grows monotonically with ITD. The best delay does not have the function of compensating for the external delay. Different realizations of this concept are possible. The most direct realization is to compare the summed activity in the two hemispheres, including the few neurons with atypical best delay (See Fig. 2C). In another version, the summed activity of all neurons with positive best delay can be compared to those with negative best delay, with the difficulty of assigning a hemisphere to neurons with a best delay near zero. In practice, there is little difference between these since the sign of the best delay is largely shared by neurons in a single hemisphere.

A third concept involves considering the total activity of all of the coincidence-counting neurons as a function of their best delay, and computing the centroid (or center of mass) along the best-delay axis (Fig. 2B; e.g. Blauert and Cobben, 1978; Stern and Colburn, 1978). This concept can be applied to broad distributions of best delays and also to sparse best-delay distributions. For many signals with reinforcing ITDs of small magnitude, the best-delay centroid and the place-code mechanism provide the same information.

While the above-mentioned direction or ITD decoder are elegant analytical concepts, it can be assumed that the actual mechanism is likely neither of the above in a strict sense. It has previously been argued that neither a strict rate nor place code is in line with human perception (see, e.g., Dietz et al., 2009; Goodman et al., 2013). Several models propose a population decoder that employs information from an ensemble of differently-tuned neurons (Fig. 2D; e.g., Colburn and Isabelle, 1992; Day et al., 2013; Goodman et al., 2013).

Note that neither of the theoretical concepts claims to be a realistic model of spatial representation at cortical levels. However, they are interesting to compare with recordings from the auditory cortex: For instance the spatially broad tuning of auditory cortex neurons (Middlebrooks et al., 1994) could be considered as an argument against a strict place code (concept 1). However, recordings from multiple pathway stages (Yao et al., 2015) reveal the complexity of the transformation of the neural code. Similarly, cooling studies that allow for deactivating very specific brain areas (e.g., Malhotra and Lomber, 2007) give important insights of cortical specialization and its behavioral relevance but at the same time reveal a high level of complexity. Models have only recently started to take cortical electrophysiologic data into account (Dong et al., 2016).

In summary, understanding and realistically modelling the transformation of spatial coding is among the most challenging endeavors of binaural hearing research. Nevertheless, it appears worthwhile to compare the abilities of the existing direction or ITD decoder mechanisms in predicting the wealth of experimental data.

2.3 Interaural group delay

This section introduces and explains the acoustic signal property “interaural group delay” and then reviews four different concepts of how this is exploited by the auditory system.

Pure tones provide only a single temporal disparity, namely the interaural phase difference (IPD) at their singular frequency. The periodic nature of both the pure tone and the IPD causes ambiguities, e.g. at 1 kHz where the cycle duration is 1 ms: $ITD = +0.5$ ms is identical to $ITD = -0.5$ ms, both corresponding to $IPD = \pi$. For complex stimuli with more than one frequency component, alternative strategies are available to extract the ITD and to disambiguate IPDs which differ for the different frequency components. The change of IPD across frequency is called interaural group delay ($d\text{IPD} / d f$). Interaural group delay is an

additional physical cue complementing the IPDs. It is used to describe interaural cues in acoustic waveforms at the ears prior to processing by the auditory system (Macaulay et al., 2017).

In contrast to a pure tone IPD, it is less clear how interaural group delay is assessed by the binaural system (Dietz et al., 2012). A stimulus specifically designed to address this question is narrowband noise centered at 500 Hz with ITD = 1.5 ms (Stern et al., 1988). At the center frequency, this ITD corresponds to an IPD = $+3\pi/2$ which is equal to IPD = $-\pi/2$. Because the smaller IPD magnitude is more plausible, in case of pure tones and for very narrowband noise (which Stern et al., 1988, refer to as “centrality”), this stimulus is perceived from the side where the signal is actually lagging. As the bandwidth increases, the location of the perceptual image changes to the side that is leading in time (Stern et al., 1988; Stern and Trahiotis, 1997, 1992; Trahiotis and Stern, 1989). Several very different mechanisms have been proposed in binaural modelling studies to explain this bandwidth-dependent lateralization. All model approaches start with some form of auditory filterbank, to decompose the signal into different frequency bands and derive either an IPD or a complete cross-correlation function in each frequency band. From there on four different concepts exist:

(1) Closely related to the physical interaural group delay cue, the auditory system may determine the IPD and its change over frequency ($d \text{ IPD} / d f$). To our knowledge this concept is not directly employed in any model. Some physiologically-motivated models (e.g., Takanen et al., 2014) employ processing that effectively operates with a similar mechanism. The concept of interaural group delay was more extensively discussed by Dietz *et al.* (2012) and is in line with the physiological observation of frequency-dependent and small best delays (McAlpine et al., 2001) referred to as the π -limit (Marquardt and McAlpine, 2006).

(2) An across-frequency comparison of the cross-correlation function that emphasizes internal delays where there are peaks that are consistent over frequency. Various mechanisms have been proposed that might mediate this processing, including (a) the original black-box approach of Stern et al. (1988), (b) simple integration over frequency, but with a broader density function describing the distribution of internal delays (Shackleton et al., 1992), and (c) a second, more central, mechanism that records coincidences at the same internal delay over a range of frequencies (Stern and Trahiotis, 1997, 1992). We also note that the side at which the cross-correlation peaks align better across frequencies is also the side which the interaural group delay indicates. The alignment of peaks across frequencies is referred to as “straightness.” The difference between the straightness concept and processing based on stimulus group delay is that the group delay approach focuses on the most central peak of the cross-correlation (*i.e.* the IPD), or more precisely the tilt ($d \text{ IPD} / d f$). In contrast, with straightness weighting, neurons with larger best delays are employed, in order to allow for an across-frequency comparison at constant delays.

Similarly, the population decoders described in Sec. 2.2 (Goodman et al., 2013; Day and Delgutte, 2013) will implicitly learn a “straightness-like” cue (following Concept 2 above) or a tilt pattern (following Concept 1).

While the above two concepts operate across auditory filter bands, the third and fourth concepts extract the within-channel group delay. Another way to describe within-channel group delay is the ITD between left and right modulation, typically called envelope ITD. A good example is that for a given ITD the lateralization of a low-frequency stimulus depends on its amplitude modulation (Bernstein and Trahiotis, 1985). For small modulation frequencies such as 25 Hz this can be attributed as a within-channel cue. The two possible mechanisms to extract this cue are:

(3): Cross-correlation models with long delay lines exploit the different heights of the lobes of the cross correlation function within one filter band, without the need for across-frequency straightness weighting. This approach requires best delays that operate beyond the π -limit.

(4) The modulation components are demodulated by cochlear nonlinearities and become physically present in the modulation spectrum. This mechanism is standard for models of high-frequency binaural processing (e.g., Bernstein and Trahiotis, 2002), but is typically not used at low-frequencies, with only a few exceptions (Dietz et al., 2009; Takanen et al., 2014; Wang et al., 2014).

Both concept (3) and (4) have no across-frequency interaction, so that models strictly following one of these concepts are not expected to account for the lateralization changes of the 1.5-ms delayed noise when increasing the bandwidth from 200 to 400 Hz (e.g., Dietz et al., 2009). Takanen et al. (2014) addressed this problem by including a wide-band pathway into their model.

Models for interaural group delay processing have previously been compared, but mostly in a qualitative manner (e.g., Dietz et al., 2012; Thompson et al., 2006).

2.4 Coincidence detection

Binaural models differ wildly in how they “perform” the binaural interaction. Phenomenological models typically operate on a continuous waveform, representing something like a spike probability, rather than operating on discrete action potentials or membrane potentials. Irrespective of the questions on best delay or delay lines, the actual input/output function of a binaural neuron is often modelled by a multiplication (e.g., Colburn, 1977; Stern and Colburn, 1978), summation (van der Heijden et al., 2013), or subtraction (e.g., Breebaart et al., 2001a; Durlach, 1963). In the Jeffress-Colburn model the multiplication occurs because the probability of a coincidence in firing between two fibers is proportional to the product of the individual probabilities if responses from the two peripheral fibers are statistically independent. In addition, the function describing the relative number of coincidences in the Jeffress-Colburn model can be considered to approximate the interaural cross-correlation of the binaural signals to the two ears after the effects of the peripheral filtering, nonlinearities, and other phenomena at the level of the fibers of the auditory nerve. However, when operating on a continuous waveform it is difficult to mimic the neural signal with such simple operations, because the timing of input action potentials is not considered. For example, in all studies on envelope ITD sensitivity typically performed at carrier frequencies above 2 kHz (e.g., Bernstein and Trahiotis, 2002; Griffin et al., 2005; Klein-Hennig et al., 2011), only the initial spikes of a modulation cycle carry the time information whereas the timing of subsequent responses may not carry ITD information anymore (Dietz et al., 2016). Waveform-based models therefore have the tendency to give too much weight to the ongoing segment of a modulation cycle. Including an adaptation stage reduces the weight of the ongoing segment but does not eliminate its influence on the waveform-based cross-correlation. This is in contradiction to what is seen behaviorally (Klein-Hennig et al., 2011), in actual neurons (Greenberg et al., 2017), or in Hodgkin-Huxley type model neurons (Dietz et al., 2016). This topic appears to be well suited and timely for comparative studies, as several datasets were published in recent years (Buell et al., 2008; Dietz et al., 2016, 2015, 2014, 2013b; Joris et al., 1994; Klein-Hennig et al., 2011; Stecker and Bibee, 2014), many of which included models. Furthermore, neither the biophysical binaural models (e.g., Cai et al., 1998; Zhou et al., 2005) nor the more recent spiking phenomenological models (e.g., Ashida et al., 2016; Takanen et al., 2014) have been compared with continuous waveform-based models on the datasets referenced above.

Among the phenomenological models that operate on subtraction is the highly successful equalization-cancellation (EC) model. It was first proposed by Kock (1950) and later developed extensively by (Durlach, 1972, 1963). As originally applied to binaural detection experiments (Colburn and Durlach, 1978), the EC

model operates in two steps. First, using *a priori* knowledge about the masker ITD and ILD, the masker components of the left and right signal are equalized by imposing a compensatory time delay and level difference. With respect to the underlying binaural processing, this concept is related to the compensatory delay lines used in the place code concept (Sec. 2.2). In the second step the two channels are subtracted from each other, cancelling out the masker. A unique feature of the EC model is that the format of the output is still a waveform representation of the target. If the equalization is perfect and the target and masker arrive in a different interaural configuration, the result of the cancellation would be an output that contains the target but not the masker. Performance in the EC model is constrained by limiting the types of operations that are permitted in the equalization operation and by adding “jitter” factors which limit the precision of the equalization operations. Its predictions agree with the results of most BMLD experiments (Durlach, 1972; Domnitz and Colburn, 1976). It is also successful in accounting for the details of Huggins pitch (Durlach, 1962), binaural edge pitch (Klein and Hartmann, 1981) and binaural coherence edge pitch (Hartmann and McMillon, 2001), among other phenomena. It has also motivated the development of other related models (*e.g.* Breebaart et al., 2001a).

EC-type models differ fundamentally from models detecting interaural decorrelation or ITD/ILD fluctuations (Goupell and Hartmann, 2007), because, as described above, they extract the signal rather than just detecting its presence. Despite psychoacoustical experimental results that allow for disambiguating the concepts (Krumbholz et al., 2009), a critical direct model comparison of these types of models remains to be performed.

2.5 Binaural speech intelligibility

Another aspect where perception and physiology have not been brought into good congruence concerns the binaural speech intelligibility level difference. Similar to-tone-in noise detection, diotically-presented speech becomes more intelligible when presented with an out-of-phase masker, compared to a diotic masker (Licklider, 1948). The perceptual data are commonly modeled using EC-type models (see Sec 2.4) because, as described above, the output format is a modified waveform representation of the target. However, up to now it has not been possible to identify the neural circuits required for this concept. Conversely, physiologically-realistic models do not describe experimental measurements of speech intelligibility as of yet.

Both experimentally and with respect to modelling efforts, the focus of models of speech intelligibility has shifted from artificial pure time or phase manipulations (Licklider, 1948), to investigating spatial release from masking (SRM). SRM refers to the difference of the speech reception threshold when comparing a co-located condition with a spatially-separated talker and interferer condition. SRM is modelled by deriving the speech intelligibility index (SII) from the output of the EC stage (Beutelmann and Brand, 2006; Culling, 2007). This model concept requires not only *a priori* information about the interaural differences of the masker but also requires a clean representation of the target without any masker. More recent variants of binaural SII based models require only limited *a priori* knowledge (Cosentino et al., 2014; Mi and Colburn, 2016).

In a recent study Josupeit and Hohmann (2017) predict SRM with automatic speech recognition results rather than simply deriving the SII. They use binaural processing only to identify ITDs and ILDs of so-called glimpses that represent perceptually important time-frequency elements. They show that spatially segregating such glimpses and then grouping the information from glimpses into objects quantitatively predicts SRM of normal hearing listeners. Their word recognizer operated on the grouped glimpses, which can come from the left ear and the right ear at different time-frequency bins, but not from the output of a binaural interaction stage. Future comparative work must reveal if the so called true-binaural advantage

of SRM really requires the same type of binaural processing that underlies the BMLD, or if the binaural system is primarily helpful for source segregation (Josupeit and Hohmann, 2017; Schoenmaker et al., 2016).

3. Difficulties in comparing models

In auditory modelling, comparative studies often contrast the predictions of pairs of models or multiple variants of a single model. A very few exceptions exist where a wider range of very different model concepts are considered (*e.g.*, models of basilar membrane motion, Saremi et al., 2016). Difficulties frequently arise from the lack of available model code and/or experimental data, as well as from some discipline-specific obstacles which will be reviewed in this section. Examples will often be from binaural models but most of the issues we discuss in this section are more general in nature.

First, the vast majority of comparative studies have been carried out by model developers which include researchers that were involved in the development of at least one of the models in the comparison. There is an inevitable risk of bias in these studies. Only a few studies actually have attempted to disconnect the comparative study from the focus of their own model development and conclude that a model by a third party is more accurate for the test conditions considered (again, Saremi *et al.* 2016 is a representative example).

A second general difficulty arises from models or model concepts that exist in different specific versions in different publications. For any third party it is unclear which version of such an evolving or mutating model is best suited for an application or comparison: The latest version or the version that was used to model the most similar dataset. It cannot be assumed that models are downward compatible, *i.e.* later versions of the same model may not reproduce the results of earlier versions (because sometimes modelers do not systematically test for this).

Another difficulty for testing models in a comparative manner is that binaural models have different outputs. For example, some models must perform a complete time averaging as part of their cross-correlation (*e.g.*, Bernstein and Trahiotis, 2017), while the output of other models may be a function of time. Variations are so manifold that it is hard to find any two models that have a comparable output format. As a consequence, each model requires a very specific testing and analysis technique and the models cannot always be compared directly. In many cases they cannot even be compared easily to the experimental data.

Similarly, when modelling the outputs of specific neurons, it is beneficial from a comparative perspective that the output format of the model is identical to the output format of the experimental study. In case of extracellular recordings, this would be a vector with the spike times, rather than just a spike probability. The latter may be more elegant for specific stand-alone studies, but would not provide a direct comparison to another model which may not provide this probability.

In addition to comparing several models on the same task, an ambition of many auditory modelers is to work toward a universal model that can explain a wider range of data compared to existing models. There has been considerable success in developing multi-purpose auditory models. This also holds true for binaural models (*e.g.*, Stern and Shear, 1996; Breebaart *et al.*, 2001 a,b,c; Takanen et al., 2014, Bernstein and Trahiotis, 2017), however, none of the above-mentioned models predicts both extent of lateralization and the presence of a dichotic element. The PhD thesis of Stern (1976) is perhaps the only document that attempts to compare lateralization, discrimination, and detection phenomena in a unified quantitative manner. In part this lack of broadness is due to the use of task-specific pathway models, which are specifically serving only a certain class of experiments. Therefore, potentially unnecessary components of

a larger general model may be left out to increase computational efficiency or simplicity. However, we argue that a model modification employing another subset of auditory pathway components constitutes a new model. Even if it is based on similar concepts, as long as there is no “master pathway model” from which different components are taken without modifying the master, the models cannot be treated as one model. They could be considered as a single model if components from one pathway model would be connected to different decision stages or artificial observers, depending on the task.

It should also be noted that in some cases researchers working in parallel may develop models based on entirely different first principles that provide equivalent predictions for at least some subsets of the data. For example, Domnitz and Colburn (1976) showed that because of the mathematical characteristics of the signals involved, the EC model of Durlach (1963), the lateralization model (Haftner, 1971; Yost, 1970), the correlation (ρ) model of Osman (1973), and the auditory-nerve-based model of Colburn (1977) all predicted the same (correct) dependence of binaural detection thresholds on the interaural time and intensity differences of tonal targets in the presence of broadband maskers. Furthermore, Colburn and Durlach (1978) noted that Colburn’s auditory-nerve-based model will predict the same detection thresholds as the EC model for a broader range of targets and maskers because Colburn’s multiple coincidence-counting units may be thought of as performing the same equalization operations (at least for ITD) as the equalization stage of the EC model, with multiple “equalizations” available simultaneously in the Colburn model while the EC model is assumed to equalize only one single ITD and one single IID. The internal representation used in Colburn’s model is also the underlying representation for the position-variable model (Stern and Colburn, 1978) and many others.

A variety of further limitations exist for many models, often not because the model is conceptually incapable, but rather because attributes of the models that influence theoretical predictions such as temporal weighting of ITD sensitivity (Stecker and Brown, 2010) or overall level dependence (Dietz et al., 2013a) have not yet sparked the interest of many model developers and are thus not included in most models. In Sec. 5 below we suggest that the use of an artificial observer may simplify the treatment of some of these tasks as well as other phenomena, because even a complicated temporal weighting data analysis can and should be used “as is” from the experimental data analysis. Additionally, if several auditory pathway models are set up to work with artificial observers, they can also be compared in a very direct, fair, and easy way. This is notwithstanding that pathway models still provide different types of internal representations to the artificial observer, hence each model may require a different observer, which is still not ideal. The concept of separating a simulation into an auditory pathway model and a subsequent artificial observer will be discussed in more detail in Section 5.

4. Collections of auditory models

It is increasingly recognized that the publication of source code of a model enhances reproducibility and impact of a study (Eglen et al., 2017). In the auditory domain a variety of auditory model code collections exist and many of them are frequently used. They include different but partly overlapping models; however, their most striking difference is in their scope. This section reviews some already-existing model collections with respect to the focus of this article and highlights those that also include experimental code or data.

4.1 Models of auditory-nerve activity

Even though we primarily discuss comparisons of models of binaural interaction, the predictions of these models can be heavily influenced by the form of the (monaural) model used to characterize the processing prior to binaural interaction. Therefore, we briefly review some of the popular descriptions.

Among the first auditory-nerve simulations is the model developed by Meddis (1986), which includes a detailed physiological characterization. The model has been and is widely used as a front end for models of central auditory processing, including models of binaural interaction (e.g., Hartung and Trahiotis, 2001; Shackleton et al., 1992; Takanen et al., 2014). Over the last three decades it has been constantly extended, e.g., by adding a nonlinear filterbank (Lopez-Poveda and Meddis, 2001) or by a revised inner-hair cell model (Sumner et al., 2003). It is implemented in MATLAB (Mathworks, Inc) and currently called a “model of the auditory periphery.” In its most recent version it includes brainstem processing of various types of neurons, including binaurally sensitive neurons (Meddis 2017, this issue).

The Auditory Toolbox (Slaney, 1998) is one of the first publicly available collections of auditory models. It was made available in MATLAB, based on an earlier toolbox published as a Mathematica notebook (Slaney, 1988). It contains MATLAB implementations of several peripheral auditory models of the 1980s, including the cochlear model and/or hair cell models of Lyon (1982), Meddis (1986), and Seneff (1988), among a number of other functions. Although the Auditory Toolbox does not contain any binaural models per se, the included peripheral processing models are used as initial stages of binaural models (e.g., Pastore and Braasch, 2015). The Toolbox includes what is perhaps the most widely-used implementation of gammatone filtering in MATLAB, based on the original formulation by Patterson et al. (1992).

In addition, Carney and her colleagues have developed a long line of realistic auditory models (e.g., Zhang et al., 2001; Zilany et al., 2014, 2009) that have also been widely distributed and used as inputs for models of binaural interaction (e.g., Dietz et al., 2016; Wang and Colburn, 2012 to name just a few). Some of the features of these models include a more detailed description of various nonlinear attributes of the response, including level-dependent amplitude compression, and a control path that varies the gain and frequency selectivity of the primary response path. These augmentations enable a more accurate description of lateral suppression, amplitude-modulated transfer functions, long-term adaptation, forward masking, and the growth of suppression as a function of input gain.

While these sophisticated peripheral models have been required to explain some binaural data, however, for models of binaural perception a highly simplified model is arguably the most commonly-used pre-processing even to date. This processing consists of a filterbank followed by a static compression, half-wave rectification and low-pass filtering (e.g., Bernstein and Trahiotis, 2002).

All of these peripheral models have been of valuable service to models that focus on binaural interaction. However, they can also provide the additional benefit of a standardized front-end for comparative studies of binaural processing.

4.2 The binaural cross-correlogram toolbox

This MATLAB toolbox, developed by M. Akeroyd while at the University of Connecticut and the University of Sussex, was developed to calculate and display various kinds of cross-correlograms. A cross-correlogram refers to a display of multiple cross-correlation functions over a range of different frequency bands. The Akeroyd toolbox was designed to facilitate comparisons of major models of the time including those described in Stern and Shear (1996), Bernstein and Trahiotis, (1996), and Shackleton et al. (1992). Options included in the Toolbox enable it to develop calculations based on different types of models for peripheral auditory-nerve activity, different assumptions concerning the nature of the binaural interaction (including normalized cross-correlation, un-normalized cross-correlation, and interaural subtraction), differing assumptions concerning the weighting of the cross-correlogram across internal delay (including the distributions of best delays proposed by Colburn, 1977; Shackleton et al., 1992; and Stern and Shear, 1996) and distributions of characteristic frequency (including that of Raatgever, 1980), as well as different means of predicting lateral position from the binaural representation (including

centroid calculation, location of the maximum peak of the cross-correlogram, and location of the peak closest to the midline).

The tight integration of the stimulus generation and display components of the Toolbox has facilitated very direct comparisons of the above-mentioned models, and its implementation in MATLAB makes it relatively easy to extend in order to incorporate the assumptions of some more recent models. The binaural cross-correlogram toolbox, with some local modification, is still being used to develop the predictions for a number of modeling studies (e.g., Bernstein and Trahiotis, 2014, 2012; Cornelis et al., 2014; Stecker et al., 2013). Its most widely-used version was originally released in 2001 and it has recently been made available online (Akeroyd, 2017) licensed to The University of Nottingham by the Creative Commons attribution, noncommercial, no derivatives (BY-NC-ND) license ([dx.doi.org/10.17639/nott.320](https://doi.org/10.17639/nott.320)).

4.3 The Auditory Modeling Toolbox

The Auditory Modeling Toolbox (AMT, Søndergaard and Majdak, 2013) contains one of the largest collections of auditory models. The AMT focuses on reproducing published research results and providing model implementations that can be used by other researchers. To this end, the AMT is more than just a collection: the models are implemented following a list of code requirements, allowing a comprehensive documentation and facilitating modular usage of the models.

In the currently-released Version 0.9.8, the AMT incorporates over 25 models. The AMT is a particularly rich repository of binaural models because it is the umbrella project of the group “Aural assessment by means of binaural algorithms” (AABBA, Blauert, 2013). It includes models of the auditory periphery at the levels of the outer, middle, and inner ear. Models of binaural interaction include delay-line based models with cross-correlation (Lindemann, 1986) or subtraction (Breebaart et al., 2001a), as well as models without explicit delay lines (Dietz et al., 2011; Takanen et al., 2014). In addition, various models of binaural or spatial perception extract information from the binaural interaction models such as tone-in-noise detection (Breebaart et al., 2001a; Culling et al., 2004) or horizontal sound source location (Dietz et al., 2011; May et al., 2011; Wierstorf et al., 2013). Other AMT models of auditory perception exist for distance estimation (Georganti et al., 2013), sagittal plane localization (Baumgartner et al., 2016) as well as multiple models for speech perception.

The AMT is licensed by the general public license (GPL), and hence is a free open-source toolbox (amtoolbox.sourceforge.net). The core of the AMT is implemented in MATLAB. Although the AMT also incorporates models implemented in Python, C, and C++, it presents itself to the user as a MATLAB/Octave toolbox.

In the AMT, the models are typically contributed by the original model developer. Because of the coding conventions and documentation rules, the AMT presents a moderately difficult entry level for model contributors.

The AMT has facilitated comparative model studies in the past (e.g., Saremi et al., 2016). Particularly helpful for comparative modelling is the collection of various models that are specific to the various pathway stages that have been mentioned above. Within the toolbox these stage-specific models are assessed by the multi-stage models that typically produce an internal representation or other output of various kinds.

Despite the immense value of reproducing results from a single modelling study, most of the multi-stage models within AMT have a specific output that differs from model to model. Therefore it remains difficult to compare the results of different models directly.

4.4 EarLab

EarLab was developed mainly in the years of 2001-2006 by David Mountain at the Boston University Hearing Research Center (Mountain et al., 2005). While its neuroscientific goals were primarily to develop integrated models of sound-source localization, it is mostly used for modelling cochlear and auditory-nerve processing (e.g., Wang and Colburn, 2012).

In addition, the EarLab Project had a strong focus on overcoming the module-connectivity problems described in Sec. 3. The EarLab Project embedded a large number of auditory models of specific auditory pathway stages in an elaborate software framework that allows for a flexible combination of modules. In contrast to the Auditory Modeling Toolbox, the modules were re-implemented by the project members in order to guarantee connectivity with modules from other stages. This high level of module integration incurs a very high cost of entry. It is not easily possible to adapt an existing model to work as an EarLab module.

It had been possible to operate EarLab over a web interface, but this and some other Earlab services have been discontinued. Most of the modules can be downloaded from the project website (earlab.bu.edu).

4.5 The Two!Ears toolbox

The Two!Ears Toolbox originates from a European multi-partner project with the overall goal “to develop an intelligent, active computational model of auditory perception and experience in a multi-modal context” (twoears.eu). Similar to EarLab, Two!Ears is not directly an open collection of auditory models, but rather a managed collection of specific models that are thought to serve the project goal. The toolbox has a specific focus on binaural hearing – less on the detailed aspects of binaural processing in the brainstem, but rather on the entire human spatial perception of sounds. For the peripheral signal processing, it relies on the models implemented in the AMT. In addition, it prominently features the concept of an artificial observer by employing complex learning and hypothesis building techniques, including the ability to rotate the listener’s head (Ma and Brown, 2016). In order to allow for such functionality, all components in the Two!Ears toolbox are capable of real-time processing of incoming binaural signals.

Some of the Two!Ears toolbox features are focused on the application of binaural signal processing in devices based on artificial-intelligence and human-machine interactions, and some of the features involve machine-learning algorithms such as deep neural networks. While this is reasonable in order to enable functionality in applications, analysis based on deep learning is currently beyond the scope of most binaural model developers focusing on understanding and modelling the mechanisms of the binaural nuclei in the brain.

The Two!Ears peripheral model is not a single pathway model but, similar to the other toolboxes, a collection of exchangeable models for each stage. Among the toolboxes described here, the Two!Ears toolbox has the highest level of module integration, with components that relate to model comparison and testing. Inevitably it is thus more challenging for an external contributor to adapt existing model code to the detailed requirements of this toolbox.

4.6 Brian Hears

“Brian Hears” is a software package designed to facilitate the creation of novel auditory models (Fontaine et al., 2011). Rather than focusing on providing complete implementations of existing models, it provides modular and configurable components that are commonly used in many auditory models, reducing the time and effort required to create a new model. The focus is on creating filterbanks to model the

peripheral auditory system. To model more central auditory neurons, the package is designed to integrate easily with the general purpose “Brian” spiking neural network simulator (Goodman and Brette, 2008, 2009).

Brian Hears includes a library of passive filterbanks such as the gammatone (Patterson et al., 1992), as well as facilities for creating active filterbanks with feedback from potentially multiple pathways. It also includes code for using head-related transfer functions (HRTFs). Technically, Brian Hears works on buffered segments of audio rather than keeping the entire signal and the outputs of all filterbanks in memory simultaneously, and is therefore able to work on large numbers of channels simultaneously with long sounds. For example, it is feasible and computationally efficient to implement a binaural model with the 6,000 peripheral channels and around 60,000 auditory nerve fibers found in the human auditory system.

Brian Hears does not include any explicit features that specifically facilitate model comparison. However, all models created using the package work on arbitrary input waveforms, and it is therefore straightforward to include them in a battery of tests and comparisons. As is true for most of the other toolboxes, it is also open source and freely available (briansimulator.org/docs/hears.html).

4.7 The AFC Framework

AFC (which is an acronym for “alternative forced choice”) is a modular framework for running psychoacoustical experiments and computational perception models (Ewert, 2013; aforcedchoice.com) in MATLAB and Octave. It is freely available for non-commercial use. While its most frequent use is for psychoacoustical experiments with human listeners, AFC was also designed to interface with perceptual models to perform psychoacoustical experiments in exactly the same framework as the measurements with human listeners. Example models are provided in the release. AFC has been extensively used (e.g., Jepsen et al., 2008) to run artificial-observer-based (monaural) models based on Dau et al. (1997a, 1997b). As previously discussed, this functionality is beneficial, if not essential, for a fair comparison of multiple perceptual models. The integration of experiment and model framework into one program provides a seamless interface but requires the model I/O to be adapted to the AFC conventions. Thus AFC offers everything needed for comparing several models to experimental data obtained with AFC; however, it is not possible to develop predictions for models on other experiments.

5. An approach to facilitate direct comparison of model components

One conclusion from the previous section is that while all of these toolboxes have been helpful for some comparative studies in the past, none of them is tailored to the specific needs of comparative model studies or for optimizing the testability of models. Therefore, within this section, we propose and describe a framework specific for these purposes. The required modules for model testing and comparing are described both in general terms and specific to the proposed Auditory Model Framework. Implementation details, interface specifications, and a beta version of a MATLAB and a Python implementation, including example models can be found in the supplemental material (and here: <http://github.com/model-initiative>). The framework is not considered to be a model collection at this stage, but may develop that way. Right now it includes documentation about how to integrate many of the abovementioned model and experiment collections. As example code to experience the operation of the Auditory Model Framework, it further comprises a stand-alone example with all components described below.

Figure 3 shows the schematics of the module connections and information flow for two examples: For a model of a typical neural response experiment and for an adaptive 3-interval psychoacoustical experiment with model.

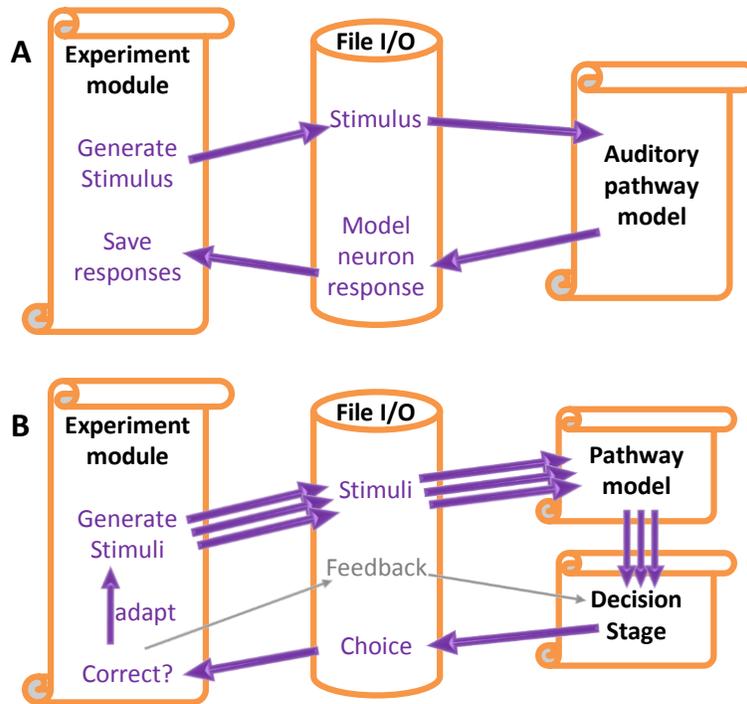


Figure 3. Two examples of the Auditory Model Framework. A: Example for models of neural response properties. B: Example for models of a psychoacoustical adaptive three-interval forced choice experiment as provided with the code package.

5.1 Measurement software routines

In order to facilitate the concept of an artificial observer described above that tests a model in the same way a subject was tested, it is best to connect the model with the measurement software routine that was used for the experiment. While the experimental description in a methods section is often more complete than a complex model description, sometimes details on stimulus generation or experimental procedure are missing – the initial value in an adaptive procedure being a prominent example. Comparability is enabled by using the same stimulation software for experiment and model.

In practice, the measurement software routine (“experiment module”) must be modified to save the sound electronically instead of sending it to the audio interface. In the current interface specifications the common Waveform Audio File Format (RIFF WAVE, known by its filename extension “.wav”) was chosen. This format also communicates implicitly the sampling frequency to the model.

All sound presentations, except for the first presentation, should be started after a “model and observer are finished” trigger. In our implementation this trigger is a file which contains the response of the artificial observer that has the same format as a real subject response. The measurement software routine can save and analyze this response.

The experiment module should also provide feedback to a model, if feedback is provided to the real subject. To date binaural models do not exploit such feedback but smart future artificial observers may require this information. Finally, some models may require a priori information, even though it is generally better to avoid this channel of information.

5.2 Auditory pathway models

There is no universally-accepted definition of where the auditory pathway starts. Typically at some level of the outer ear or at the tympanic membrane. Prior to auditory processing there is acoustic propagation which transforms a single or multi-channel speaker or headphone output to the 2-channel “binaural” input of the auditory system.

Accordingly, our stage of the auditory pathway begins with a 2-channel waveform representing the stimulus arriving at the beginning of the left and right auditory pathway of the subject that is modelled. The exact starting point has to be defined for each specific model. If it is not equal with the headphone position or the source is not a headphone-like device, the simulated experimental procedure requires a description of the acoustic propagation paths (e.g., head-related transfer functions, headphone transfer functions, or the full sound fields produced by loudspeaker-based systems). Note further that some pathway models incorporate the transfer function of the middle ears, but, particularly in binaural research, others do not. The synthesis of the binaural signals must consider such configurations.

The output of auditory pathway model is typically referred to as an internal representation. It can vary from the neural response as a function of time to a single scalar value, such as the time and frequency and delay-averaged centroid of a cross-correlogram as in Stern and Shear (1996). Figure 4 displays some exemplary internal representations of five different binaural models to the same three dichotic stimuli. However, these simple visualizations can only plot an arbitrary one- or two-dimensional selection of the underlying, typically richer, internal representation. Such plots of pathway model outputs can potentially be misleading about what information the complete internal representation contains. If care is taken and no relevant output is discarded in the display, it can nicely illustrate conceptual differences of how certain cues are represented by different models, which is relevant for the selection of appropriate decision stages.

Most auditory pathway models to date are purely signal driven. This means that the output does not depend on the state of attention, or on any long-term adaptation, such as the statistical distribution of test stimuli, despite experimental evidence that such influences exist (e.g., Maier et al., 2012). Therefore, for comparing existing auditory pathway models, it is currently not necessary to provide input of previous stimuli or attention, but it may be necessary at some point in the future. For some models additional a priori knowledge on the stimulus is required. For example, Bernstein and Trahiotis’s implementation of the position-variable model requires knowledge of the modulation frequency if it operates on high-carrier-frequency amplitude-modulated tones (Bernstein and Trahiotis, 2012, 2003).

Finally, computational constraints may impose model-specific accommodation. Instead of running the pathway model as described above it may be necessary instead to pre-process computationally expensive stages for a set of stimuli and then load these data based on stimulus parameters (e.g., Brughera et al., 2013).

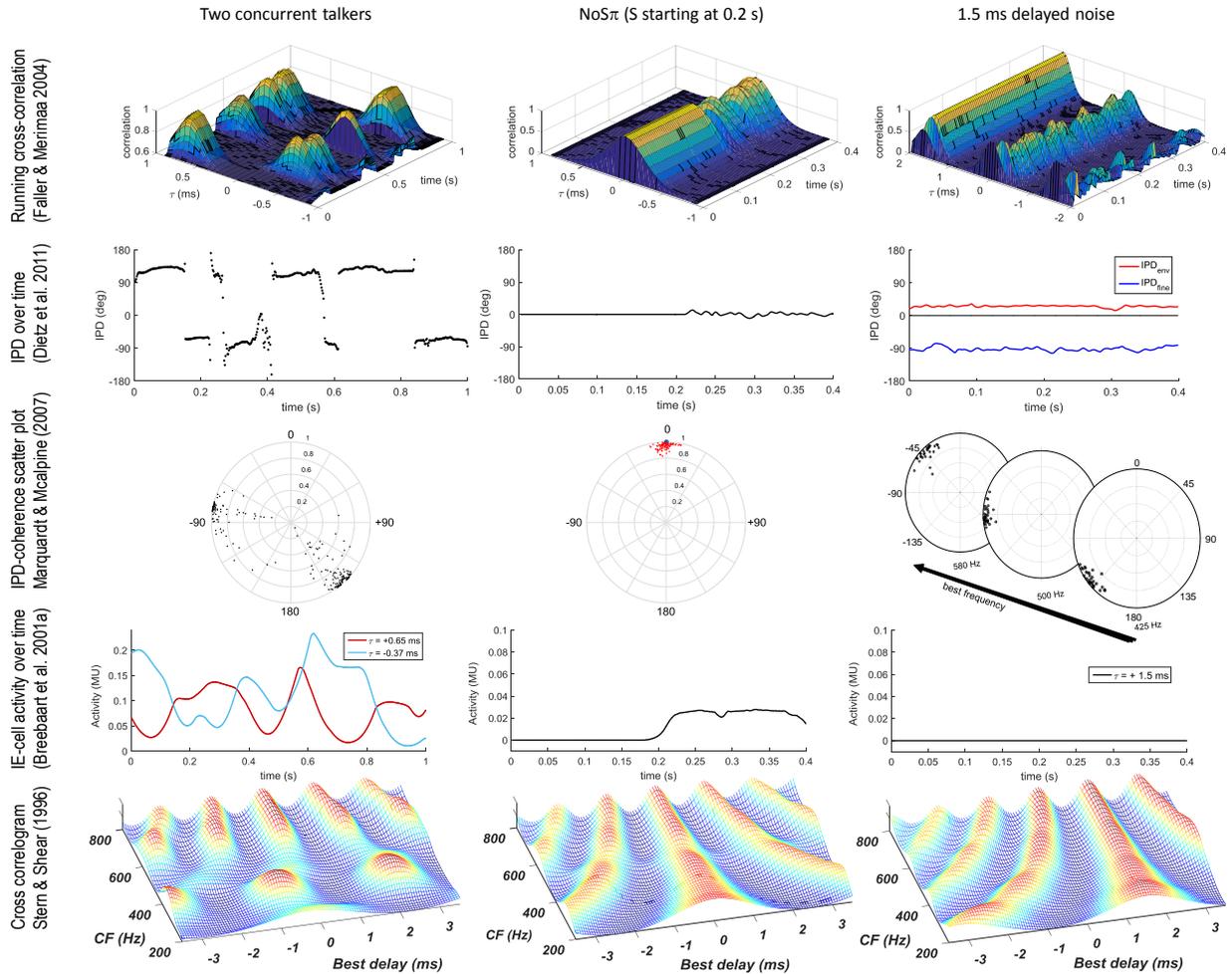


Figure 4. Exemplary illustrations of internal representations of five different models (rows) to three different stimuli (columns). Column A (left): Two competing talkers with different ITDs. Column B (middle): Diotic noise of 300 to 700 Hz. After 200 ms a 500-Hz tone is added out of phase at a signal-to-noise ratio of -15 dB. Column C: Same noise as in (B) but with an ITD = 1.5 ms. Representation 1 (top row): Running Cross-correlation (Faller and Merimaa, 2004) extracted from filter band centered at 500 Hz. Representations 2 and 3: IPD-based models. Row 2: display as in Dietz et al. 2009, 2011): IPD over time. Only the 500-Hz filter band is shown. Note that for the 1.5 ms delayed noise the envelope IPD is extracted with a 50-Hz modulation filter, so that +27 deg correspond to 1.5 ms ITD. Row 3: The same internal representation but displayed as an IPD - coherence scatter plot (Marquardt and McAlpine, 2007). In the middle panel all data points from the diotic segment are identical and indicated by the thicker blue dot. Row 4 (bottom left and middle): Time functions of outputs of selected EI-elements from the model of Breebaart et al. (2001a). For the two competing talkers two selected EI-elements with characteristic delay equal to the ITDs corresponding to the talker from the left (blue line) and from the right (red line) are shown. A high output indicates that the talker with non-corresponding ITD provides much energy at this moment. 1.5 ms delayed noise would result in constant zero activity for an EI cell with 1.5 ms best delay, which is not shown. Row 5: Cross-correlograms (Stern and Shear, 1996) generated with Akeroyd's binaural model toolbox. In this row the response is plotted as a function of characteristic frequency and best delay, averaging over time, with the signals in Columns A and B slightly modified. Column A: Response to two simultaneously-presented vowels with differing ITDs. /AH/ is presented with a -0.75 ms ITD, with a major formant at 350 Hz. /IH/ is presented with zero ITD, with major formants at 600 and 800 Hz. Column B: Response to only the NOS π signal. The presence of the target in the NOS π signal is indicated by the notch in activity at the target frequency, 500 Hz. Column C: Response to broadband noise with ITD 1.5 ms. The noise ITD is indicated by the straight ridge at 1.5 ms.

5.3 Decision stages

As mentioned in the introduction, many modelling studies focus on understanding processing along the auditory pathway. However, in order to account for behavioral data, they typically include a decision stage, which is often not in the focus and not necessarily considered to be part of the auditory pathway. It becomes particularly clear in multiple interval paradigms, where the (artificial) observer is asked to select a target interval (e.g., Breebaart et al. 2001 b,c). For both real and artificial observers, the selection of a target interval includes the comparison across sequential stimuli based on a decision-making strategy that is not necessarily attributed to signal processing in the auditory pathway. Even neurophysiological studies sometimes use a detection stage to relate neural data to behavioral data (Jiang et al., 1997; Skottun et al., 2001), or they use other means of information theory such as the Fisher information criterion to post-process neural data (e.g., Fischer and Konishi, 2008).

A decision stage is considered as a function, which takes the output from the pathway model of all intervals and provides a response identical to the real observer, e.g., the selection of a target interval or the direction of a sound source.

In behavioral experiments, feedback plays a role to optimize the response. In the future it may be possible to train artificial observers in the same way as test subjects using the feedback provided by the experimenter. Most naturally such feedback would be an additional optional input to the decision stage. Further, the decision stage should even be allowed to record past feedback and past decisions (e.g., Breebaart et al. 2001a). As with the pathway stage, some decision stages may need specific stimulus information or parameters to be set as input.

Unfortunately, there is not a single procedure to define and implement an artificial observer, especially because its operation depends on the type of internal representation provided by the pathway model. Thus, commonly, artificial observers are developed and tested simultaneously with the pathway model (e.g., Breebaart et al. 2001 a,b,c).

In auditory research, four tasks cover most psychoacoustical experiments, as well as neurometric studies which relate neural responses to perception (Jiang et al., 1997; Skottun et al., 2001)

1. Detecting a target sound in the presence of an interferer. In particular, this type includes tone-in-noise detection, with binaural masking level difference (BMLD) experiments as the most prominent binaural test.
2. Discriminating between different intervals. In binaural hearing this includes the popular ITD or ILD or direction-of-sound-arrival discrimination task (e.g., Zwillocki and Feldman, 1956). Instructions are either an odd-one-out concept, requiring at least 3 intervals, or a smaller / larger type of instruction for a specific detection cue description, often in a 2-interval procedure. A typical binaural example for the latter is “select the interval where the sound came from the right relative to the other interval” (e.g., Zwillocki and Feldman, 1956).
3. Quantitative judgement, categorization, or scaling. This classic and widely used category includes many single interval tasks, for example loudness scaling (Allen et al., 1990; Brand and Hohmann, 2002), localization (Sandel, 1955), extent of lateralization (Teas, 1962; Bernstein and Trahiotis, 1985), and sound-quality-related aspects (Fastl, 2006).
4. Predicting the change in speech intelligibility due to different ITDs and/or ILDs of target and interferer.

A single pathway model should in theory be able to serve very different artificial observers for the above categories. It is not expected that all models can serve all tasks. For example some models include a time

averaging within their binaural interaction and will not be able to operate on tasks that require time-dependent internal representations. That said and referring back to Sec. 3, there is merit to using the exact same pathway model for different tasks, even for different task categories when possible. It may cause computational overhead but it is typically a prerequisite to claim it as one general model, rather than two sister models, with highly-related processing.

The importance of employing a multi-purpose pathway model cannot be underestimated. There are a large number of experiments that are not covered by many models, simply because the models restrict themselves already in the pathway to only the components necessary for the tasks of interest to the modelers. For example, one of the most successful model concepts for predicting ITD-detection thresholds, the normalized cross-correlation coefficient model of Bernstein and Trahiotis (2009) only provides predictions at a reference ITD of zero microseconds. It is noteworthy that the underlying pathway model of Stern and Shear (1996) did include a full set of internal delays but the normalized cross-correlation coefficient model used by Bernstein and Trahiotis was simplified to restrict itself to considerations of internal delays of 0. This restriction could be relaxed easily, though.

5.4 The experiment – model interface

The last module of the framework that is required for a fair and easy comparison between real observers and a computer model, or between several computer models, is an interface connecting the three modules mentioned above.

A practical issue which may have limited the use of artificial observers in the past is the connection of the experimental operating environment with a computer model. Especially if one or both modules were developed by different third parties, this task may be difficult or impossible. Modules may even operate in different computer languages. For example, many experiments are conducted using MATLAB (e.g. Ewert, 2013), whereas models are written in either MATLAB (see e.g., Søndergaard and Majdak, 2013 for a collection) or Python (e.g., Goodman and Brette, 2009; Verhulst et al., 2012).

An exemplary solution that is used in our implementation is to communicate via electronic file exchange. The experiment module writes the sound output to wav-files and can provide additional files with a priori information and feedback. A model server reads the files produced by the experiment module and handles the model and decision stage I/O. It writes the decision in a response file.

The advantage of the file interface is that it is independent of the software that is used for each of the other modules, because file I/O is a fairly universal functionality. As our implementation is providing a model server script, the modelers do not have to deal with the file I/O; they can directly run their pathway model and any compatible decision stage that produces output in the same format and unit as would be produced by a real subject. Inclusion of an experiment module in the framework requires the functionality to write the stimulus intervals as .wav files and to read a response from a file.

5.5 Experimental data

The framework presented above implies the availability of useful experimental data, against which the models can be compared.

Data are obviously published but often in a highly processed form. In order to be fully informative, the post-processing scripts used for the experimental data should be available and be applicable to the simulated data. Furthermore, numerical availability of raw or processed data in electronic form is helpful to quantitatively compare it with model predictions. Fortunately an increasing number of experimentalists sees these advantages and provides suitable data in supplemental material. Some journals (e.g. PLOS ONE)

have updated their policies and make data availability a prerequisite. In contrast, the older literature contains several examples of highly referenced and informative studies that are in fact self-described as “preliminary” (e.g., Klumpp and Eady, 1956; Zwislocki and Feldman, 1956) or include informal but easily replicated self-reports (Trahiotis and Stern, 1994). Depending on the specific study of interest it may thus be decided by modelers that experiments need to be repeated in order to provide the same data format and to fulfill the same standards that we expect from present modelling studies. Our framework provides the infrastructure for such experiments as well. Selecting measurement software (Sec. 5.1) that is compatible with the framework will ensure that the experiments can also be conducted by artificial observers.

The framework has a folder specifically for data, a folder specifically for data analysis scripts, and a folder for plotting routines. All these folders with their functions and scripts are intended to be used interchangeably for experimental data and data obtained from an artificial observer.

6. Summary and thoughts on the future

In summary, we have described a number of obstacles to comparing binaural models quantitatively, as well as to test the predictions of third party models on own experiments. We also briefly discussed selected previous comparative studies. Echoing the conclusions of several articles (Thompson et al. 2006; Dietz et al. 2012; Leibold and Grothe 2015) there are still several conceptual conflicts both within and across the two main domains of experimental binaural research, i.e. between physiology and psychoacoustics. Binaural models and in particular quantitative comparative approaches are of growing importance to help resolving these conflicts.

Speaking for us, we have not sufficiently prioritized the applicability and comparability of the models that we have developed in the past (Dietz et al., 2009; Ewert et al., 2010; Goodman et al., 2013; Hartmann et al., 2016; Majdak et al., 2014; Marquardt and McAlpine, 2009; Stern and Shear, 1996). We are not able to compare the majority of our models in their published form. While acknowledging the unique aims of all existing models, there is going to be added value if model developers try to provide pathway models that can operate with different artificial observers, in the same way as a subject performs the different experimental tasks. We also acknowledge that model developers, and particularly pure experimentalists, are not necessarily keen on getting several different third-party models in different programming languages running, or to change the programming of their own model to the conventions of the existing toolboxes. Reflecting on the above has defined the priority for our proposed model comparison framework: Not another integrated model toolbox but rather little or no coding conventions. The only requirement is to run the model in the exact same paradigm as it was used in the experimental study.

The framework provided in the supplemental material will allow for an easy replication of published model data by third-party scientists and it will allow experimentalists to test the models on their own stimuli and experimental method. In a best-case scenario, the framework will foster collaboration or competition within specific focus groups that work to solve one of the open questions of binaural hearing, or, with minor modifications, other branches of hearing research. Once a model comparison is performed with an initial set of experiments, new experiments can be added without the need to adapt the pathway models. The most promising, best-performing models may inspire follow-up research by a broader range of researchers, in an effort to add functionality that is not is the focus of the original model developer, or to refine the model further, or to develop experiments that specifically challenge such a model. At the optimistic end of the process the community will have refined or developed a model beyond the capabilities of a single laboratory: a very versatile pathway with binaural processing concepts which have

a fair degree of physiological plausibility, which can serve a diverse range of decision stages, and consequently, which can account for a broad range of experimental data.

Acknowledgements

We thank Laurel Carney for giving us the opportunity to contribute an article on this topic in this special issue. We further thank Volker Hohmann, Constantine Trahiotis, and Steve Colburn for a motivating and fruitful initial discussions that helped triggering this initiative, Graham Voysey for his feedback on the framework code, and Steve Colburn for multiple helpful comments on the manuscript. This research was undertaken, in part, thanks to funding from the Canada Research Chairs program.

References

- Akeroyd, M.A., 2017. A binaural cross-correlogram toolbox for MATLAB.
- Allen, J.B., Hall, J.L., Jeng, P.S., 1990. Loudness growth in 1/2-octave bands (LGOB)—A procedure for the assessment of loudness. *J. Acoust. Soc. Am.* 88, 745–753. doi:10.1121/1.399778
- Ashida, G., Kretzberg, J., Tollin, D.J., 2016. Roles for Coincidence Detection in Coding Amplitude-Modulated Sounds. *PLoS Comput. Biol.* 12, 1–27. doi:10.1371/journal.pcbi.1004997
- Batra, R., Kuwada, S., Fitzpatrick, D.C., 1997. Sensitivity to interaural temporal disparities of low- and high-frequency neurons in the superior olivary complex. I. Heterogeneity of responses. *J Neurophysiol* 78, 1222–1236.
- Baumgartner, R., Majdak, P., Laback, B., 2016. Modeling the Effects of Sensorineural Hearing Loss on Sound Localization in the Median Plane. *Trends Hear.* 20, 1–11. doi:10.1177/2331216516662003
- Bernstein, L.R., Trahiotis, C., 2017. An interaural-correlation-based approach that accounts for a wide variety of binaural detection data. *J. Acoust. Soc. Am.* 141. doi:10.1121/1.4976098
- Bernstein, L.R., Trahiotis, C., 2014. Sensitivity to envelope-based interaural delays at high frequencies: Center frequency affects the envelope rate-limitation. *J. Acoust. Soc. Am.* 135, 808–816. doi:10.1121/1.4861251doi.org/10.1121/1.4861251
- Bernstein, L.R., Trahiotis, C., 2012. Lateralization produced by interaural temporal and intensive disparities of high-frequency, raised-sine stimuli : Data and modeling. *J Acoust Soc Am* 131, 409–415. doi:10.1121/1.3662056
- Bernstein, L.R., Trahiotis, C., 2009. How sensitivity to ongoing interaural temporal disparities is affected by manipulations of temporal features of the envelopes of high-frequency stimuli. *J. Acoust. Soc. Am.* 125, 3234–42. doi:10.1121/1.3101454
- Bernstein, L.R., Trahiotis, C., 2003. Enhancing interaural-delay-based extents of laterality at high frequencies by using “transposed stimuli”. *J Acoust Soc Am* 113, 3335–3347.
- Bernstein, L.R., Trahiotis, C., 2002. Enhancing sensitivity to interaural delays at high frequencies by using “transposed stimuli”. *J Acoust Soc Am* 112, 1026–1036.
- Bernstein, L.R., Trahiotis, C., 1996. On the use of the normalized correlation as an index of interaural envelope correlation. *J Acoust Soc Am* 100, 1754–1763.
- Bernstein, L.R., Trahiotis, C., 1985. Lateralization of low-frequency, complex waveforms: the use of envelope-based temporal disparities. *J Acoust Soc Am* 77, 1868–1880.
- Beutelmann, R., Brand, T., 2006. Prediction of speech intelligibility in spatial noise and reverberation for normal-hearing and hearing-impaired listeners. *J Acoust Soc Am* 120, 331–342. doi:10.1121/1.2202888
- Blauert, J., 2013. *The Technology of Binaural Listening*. Springer, Berlin.
- Blauert, J., Cobben, W., 1978. Some consideration of binaural cross-correlation analysis. *Acustica* 39, 96–103.
- Brand, A., Behrend, O., Marquardt, T., McAlpine, D., Grothe, B., 2002. Precise inhibition is essential for microsecond interaural time difference coding. *Nature* 417, 543–547. doi:10.1038/417543a
- Brand, T., Hohmann, V., 2002. An adaptive procedure for categorical loudness scaling. *J. Acoust. Soc. Am.* 112, 1597–1604. doi:10.1121/1.1502902
- Breebaart, J., van de Par, S., Kohlrausch, A., 2001a. Binaural processing model based on contralateral inhibition. I. Model structure. *J Acoust Soc Am* 110, 1074–1088.
- Breebaart, J., van de Par, S., Kohlrausch, A., 2001b. Binaural processing model based on contralateral inhibition. II. Dependence on spectral parameters. *J Acoust Soc Am* 110, 1089–1104.
- Breebaart, J., van de Par, S., Kohlrausch, A., 2001c. Binaural processing model based on contralateral inhibition. III. Dependence on temporal parameters. *J Acoust Soc Am* 110, 1105–1117.
- Brughera, A., Dunai, L., Hartmann, W.M., 2013. Human interaural time difference thresholds for sine tones: The high-frequency limit. *J. Acoust. Soc. Am.* 133, 2839–55. doi:10.1121/1.4795778
- Buell, T.N., Griffin, S.J., Bernstein, L.R., 2008. Listeners’ sensitivity to “onset/offset” and “ongoing” interaural delays in high-frequency, sinusoidally amplitude-modulated tones. *J Acoust Soc Am* 123, 279–294. doi:10.1121/1.2816399
- Cai, H., Carney, L.H., Colburn, H.S., 1998. A model for binaural response properties of inferior colliculus neurons. I. A model with

interaural time difference-sensitive excitatory and inhibitory inputs. *J. Acoust. Soc. Am.* 103, 494–506. doi:10.1121/1.421130

Carr, C.E., Konishi, M., 1990. A circuit for detection of interaural time differences in the brain stem of the barn owl. *J Neurosci* 10, 3227–3246.

Carr, C.E., Konishi, M., 1988. Axonal delay lines for time measurement in the owl's brainstem. *Proc Natl Acad Sci U S A* 85, 8311–8315.

Cherry, C.E., 1953. Some experiments on the recognition of speech, with one and with two ears. *J Acoust Soc Am* 25, 975–979. doi:10.1121/1.1907229

Chung, Y., Delgutte, B., Colburn, A.H.S., 2015. Modeling Binaural Responses in the Auditory Brainstem to Electric Stimulation of the Auditory Nerve. *J Assoc Res Otolaryngol* 15, 135–158. doi:10.1007/s10162-014-0492-6

Colburn, H. S.; Durlach, N.I., 1978. Models of binaural interaction, in: E. C. Carterette and M. P. Friedman, E. (Ed.), *Handbook of Perception*. Vol. 4, Hearing. Academic Press, New York.

Colburn, H.S., Isabelle, S.K., 1992. Models of binaural processing based on neural patterns in the medial superior olive. In: Cazals Y, Demaney L, Horner K, editors. *Auditory physiology and perception*. Oxford: Pergamon; pp. 539–545.

Colburn, H.S., 1977. Theory of binaural interaction based on auditorynerve data. II. Detection of tones in noise. *J Acoust Soc Am* 61, 525–533.

Cornelis, B., Moonen, M., Wouters, J., 2014. Reduced-bandwidth Multi-channel Wiener Filter based binaural noise reduction and localization cue preservation in binaural hearing aids. *Signal Processing* 99, 1–16. doi:10.1016/j.sigpro.2013.12.012

Cosentino, S., Marquardt, T., McAlpine, D., Culling, J.F., Falk, T.H., 2014. A model that predicts the binaural advantage to speech intelligibility from the mixed target and interferer signals. *J. Acoust. Soc. Am.* 135, 796–807. doi:10.1121/1.4861239

Culling, J.F., 2007. Evidence specifically favoring the equalization-cancellation theory of binaural unmasking. *J Acoust Soc Am* 122, 2803–2813. doi:10.1121/1.2785035

Culling, J.F., Hawley, M.L., Litovsky, R.Y., 2004. The role of head-induced interaural time and level differences in the speech reception threshold for multiple interfering sound sources. *J Acoust Soc Am* 116, 1057–1065.

Dau, T., Kollmeier, B., Kohlrausch, A., 1997a. Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers. *J Acoust Soc Am* 102, 2892–2905. doi:10.1121/1.420344

Dau, T., Kollmeier, B., Kohlrausch, A., 1997b. Modeling auditory processing of amplitude modulation. II. Spectral and temporal integration. *J Acoust Soc Am* 102, 2906–2919. doi:10.1121/1.420344

Day, M.L., Koka, K., Delgutte, B., 2013. of a concurrent, spatially separated source Neural encoding of sound source location in the presence of a concurrent, spatially separated source. *J. Neurophysiol. Univ. Oldenburg/IBIT July J Neurophysiol* 31, 2612–2628. doi:10.1152/jn.00303.2012

Dietz, M., 2016. Models of the electrically stimulated binaural system: A review. *Network* 27, 186–211. doi:10.1080/0954898X.2016.1219411

Dietz, M., Bernstein, L.R., Trahiotis, C., Ewert, S.D., Hohmann, V., 2013a. The effect of overall level on sensitivity to interaural differences of time and level at high frequencies. *J. Acoust. Soc. Am.* 134, 494–502. doi:10.1121/1.4807827

Dietz, M., Ewert, S.D., Hohmann, V., 2012. Lateralization based on interaural differences in the second-order amplitude modulator. *J. Acoust. Soc. Am.* 131, 398–408. doi:10.1121/1.3662078

Dietz, M., Ewert, S.D., Hohmann, V., 2011. Auditory model based direction estimation of concurrent speakers from binaural signals. *Speech Commun.* 53, 592–605. doi:10.1016/j.specom.2010.05.006

Dietz, M., Ewert, S.D., Hohmann, V., 2009. Lateralization of stimuli with independent fine-structure and envelope-based temporal disparities. *J Acoust Soc Am* 125, 1622–1635. doi:10.1121/1.3076045

Dietz, M., Klein-Hennig, M., Hohmann, V., 2015. The influence of pause, attack, and decay duration of the ongoing envelope on sound lateralization. *J. Acoust. Soc. Am.* 137, EL137-43. doi:10.1121/1.4905891

Dietz, M., Marquardt, T., Salminen, N.H., McAlpine, D., 2013b. Emphasis of spatial cues in the temporal fine structure during the rising segments of amplitude-modulated sounds. *Proc. Natl. Acad. Sci. U. S. A.* 110, 15151–6. doi:10.1073/pnas.1309712110

Dietz, M., Marquardt, T., Stange, A., Pecka, M., Grothe, B., McAlpine, D., 2014. Emphasis of spatial cues in the temporal fine structure during the rising segments of amplitude-modulated sounds II: single-neuron recordings. *J. Neurophysiol.* 111, 1973–85. doi:10.1152/jn.00681.2013

Dietz, M., Wang, L., Greenberg, D., McAlpine, D., 2016. Sensitivity to Interaural Time Differences Conveyed in the Stimulus Envelope: Estimating Inputs of Binaural Neurons Through the Temporal Analysis of Spike Trains. *JARO - J. Assoc. Res. Otolaryngol.* 17, 313–330. doi:10.1007/s10162-016-0573-9

Domnitz, R., Colburn, H., 1976. Analysis of binaural detection models for dependence on interaural target parameters. *J. Acoust. Soc. Am.* 59, 598–601. doi:10.1121/1.380904

Dong, J., Colburn, H.S., Sen, K., 2016. Cortical Transformation of Spatial Processing for Solving the Cocktail Party Problem: A Computational Model. *eNeuro* 3, 1–11. doi:10.1523/ENEURO.0086-15.2015

Durlach, N.I., 1972. Binaural signal detection: Equalization and cancellation theory, in: Tobias, J. (Ed.), *Foundations of Modern Auditory Theory* Volume 2. Academic Press, New York, pp. 369–462.

Durlach, N.I., 1963. Equalization and Cancellation Theory of Binaural Masking-Level Differences. *J. Acoust. Soc. Am.* 35, 1206–1218. doi:10.1121/1.1918675

- Eglen, S., Marwick, B., Halchenko, Y., Hanke, M., Sufi, S., Gleeson, P., Silver, R.A., Davison, A., Lanyon, L., Abrams, M., Wachtler, T., Willshaw, D.J., Pouzat, C., Poline, J.-B., 2017. Towards standard practices for sharing computer code and programs in neuroscience. *bioRxiv* 20, 45104. doi:10.1101/045104
- Ewert, S.D., 2013. AFC - a modular framework for running psychoacoustic experiments and computational perception models, in: *Proceedings of the International Conference on Acoustics (AIA-DAGA 2013)*. pp. 1326–1329.
- Ewert, S.D., Dietz, M., Klein-hennig, M., 2010. The Neurophysiological Bases of Auditory Perception 337–346. doi:10.1007/978-1-4419-5686-6
- Faller, C., Merimaa, J., 2004. Source localization in complex listening situations: selection of binaural cues based on interaural coherence. *J Acoust Soc Am* 116, 3075–3089. doi:10.1121/1.1791872
- Fastl, H., 2006. Psychoacoustic basis of sound quality evaluation and sound engineering. *Proceeding ICSV13 - 13th Int. Congr. Sound Vib.* 2006 39, 324–345. doi:10.2307/2334029
- Fischer, B.J., Konishi, M., 2008. Variability reduction in interaural time difference tuning in the barn owl. *J. Neurophysiol.* 100, 708–715. doi:10.1152/jn.90358.2008
- Fontaine, B., Goodman, D.F.M., Benichoux, V., Brette, R., 2011. Brian Hears: Online Auditory Processing Using Vectorization Over Channels. *Front. Neuroinform.* 5, 1–9. doi:10.3389/fninf.2011.00009
- Georganti, E., May, T., Van De Par, S., Mourjopoulos, J., 2013. Sound source distance estimation in rooms based on statistical properties of binaural signals. *IEEE Trans. Audio, Speech Lang. Process.* 21, 1727–1741. doi:10.1109/TASL.2013.2260155
- Goodman, D., Brette, R., 2008. Brian: a simulator for spiking neural networks in Python. *BMC Neurosci.* 9, P92. doi:10.1186/1471-2202-9-S1-P92
- Goodman, D.F.M., Benichoux, V., Brette, R., 2013. Decoding neural responses to temporal cues for sound localization. *Elife* 2013, 1–21. doi:10.7554/eLife.01312
- Goodman, D.F.M., Brette, R., 2009. The brian simulator. *Front. Neurosci.* 3, 192–197. doi:10.3389/neuro.01.026.2009
- Goupell, M.J., Hartmann, W.M., 2007. Interaural fluctuations and the detection of interaural incoherence. III. Narrowband experiments and binaural models. *J Acoust Soc Am* 122, 1029–1045. doi:10.1121/1.2734489
- Greenberg, D., Monaghan, J.J.M., Dietz, M., Marquardt, T., McAlpine, D., 2017. Influence of envelope waveform on ITD sensitivity of neurons in the auditory midbrain. *J Neurophysiol* 118, 2358–2370
- Griffin, S.J., Bernstein, L.R., Ingham, N.J., McAlpine, D., 2005. Neural sensitivity to interaural envelope delays in the inferior colliculus of the guinea pig. *J Neurophysiol* 93, 3463–3478. doi:10.1152/jn.00794.2004
- Haftner, E.R., 1971. Quantitative Evaluation of a Lateralization of Masking-Level Differences Model 50, 1116–1122.
- Harper, N.S., Scott, B.H., Semple, M.N., McAlpine, D., 2014. The neural code for auditory space depends on sound frequency and head size in an optimal manner. *PLoS One* 9. doi:10.1371/journal.pone.0108154
- Hartmann, W.M., McMillon, C.D., 2001. Binaural coherence edge pitch. *J. Acoust. Soc. Am.* 109, 294–305. doi:10.1121/1.2021327
- Hartmann, W.M., Rakerd, B., Crawford, Z.D., Zhang, P.X., 2016. Transaural experiments and a revised duplex theory for the localization of low-frequency tones. *J. Acoust. Soc. Am.* 139, 968. doi:10.1121/1.4941915
- Hartung, K., Trahiotis, C., 2001. Peripheral auditory processing and investigations of the “precedence effect” which utilize successive transient stimuli. *J. Acoust. Soc. Am.* 110, 1505. doi:10.1121/1.1390339
- Jeffress, L.A., 1948. A place theory of sound localization. *J. Comp. Physiol. Psychol.* 41, 35–39.
- Jepsen, M.L., Ewert, S.D., Dau, T., 2008. A computational model of human auditory signal processing and perception. *J Acoust Soc Am* 124, 422–438. doi:10.1121/1.2924135
- Jiang, D., McAlpine, D., Palmer, A.R., 1997. Detectability index measures of binaural masking level difference across populations of inferior colliculus neurons. *J. Neurosci.* 17, 9331–9.
- Joris, P.X., Carney, L.H., Smith, P.H., Yin, T.C., 1994. Enhancement of neural synchronization in the anteroventral cochlear nucleus. I. Responses to tones at the characteristic frequency. *J. Neurophysiol.* 71, 1022–36.
- Joris, P.X., Yin, T.C.T., 2007. A matter of time: internal delays in binaural processing. *Trends Neurosci* 30, 70–78. doi:10.1016/j.tins.2006.12.004
- Josupeit, A., Hohmann, V., 2017. Modeling speech localization, talker identification, and word recognition in a multi-talker setting. *J Acoust Soc Am* in press.
- Klein-Hennig, M., Dietz, M., Hohmann, V., Ewert, S.D., 2011. The influence of different segments of the ongoing envelope on sensitivity to interaural time delays. *J. Acoust. Soc. Am.* 129, 3856–72. doi:10.1121/1.3585847
- Klein, M.A., Hartmann, W.M., 1981. Binaural edge pitch. *J. Acoust. Soc. Am.* 70, 51–61. doi:10.1121/1.386581
- Klumpp, R.G., Eady, H.R., 1956. Some measurements of interaural time difference thresholds. *J Acoust Soc Am* 28, 859–860. doi:10.1121/1.1908493
- Kock, W.E., 1950. Binaural Localization and Masking. *J. Acoust. Soc. Am.* 22, 801–804. doi:10.1121/1.1906692
- Krumbholz, K., Magezi, D.A., Moore, R.C., Patterson, R.D., 2009. Binaural sluggishness precludes temporal pitch processing based on envelope cues in conditions of binaural unmasking. *J Acoust Soc Am* 125, 1067–1074. doi:10.1121/1.3056557
- Leibold, C., Grothe, B., 2015. Sound localization with microsecond precision in mammals: what is it we do not understand? *e-Neuroforum* 6, 3–10. doi:10.1007/s13295-015-0001-3
- Licklider, J.C.R., 1948. The Influence of Interaural Phase Relations upon the Masking of Speech by White Noise. *J Acoust Soc Am*

20, 150–159.

- Lindemann, W., 1986. Extension of a binaural cross-correlation model by contralateral inhibition. I. Simulation of lateralization for stationary signals. *J Acoust Soc Am* 80, 1608–1622. doi:10.1121/1.394325
- Lopez-Poveda, E.A., Meddis, R., 2001. A human nonlinear cochlear filterbank. *J. Acoust. Soc. Am.* 110, 3107–3118. doi:10.1121/1.1416197
- Lyon, R.F., 1982. A Computational Model of Filtering, Detection, and Compression in the Cochlea, in: *Intl. Conf. on Acoust., Speech, and Sig. Proc.*, IEEE, Paris. pp. 1282–1285.
- Ma, N., Brown, G.J., 2016. Speech localisation in a multitalker mixture by humans and machines. *Proc. Annu. Conf. Int. Speech Commun. Assoc. INTERSPEECH 08–12–Sept*, 3359–3363. doi:10.21437/Interspeech.2016-1149
- Macaulay, E.J., Rakerd, B., Andrews, T.J., Hartmann, W.M., 2017. On the localization of high-frequency, sinusoidally amplitude-modulated tones in free field. *J. Acoust. Soc. Am.* 141, 847–863. doi:10.1121/1.4976047
- Maier, J.K., Hehrmann, P., Harper, N.S., Klump, G.M., Pressnitzer, D., McAlpine, D., 2012. Adaptive coding is constrained to midline locations in a spatial listening task. *J. Neurophysiol.* 108, 1856–1868. doi:10.1152/jn.00652.2011
- Majdak, P., Baumgartner, R., Laback, B., 2014. Acoustic and non-acoustic factors in modeling listener-specific performance of sagittal-plane sound localization. *Front. Psychol.* 5, 1–10. doi:10.3389/fpsyg.2014.00319
- Malhotra, S., Lomber, S.G., 2007. Sound localization during homotopic and heterotopic bilateral cooling deactivation of primary and nonprimary auditory cortical areas in the cat. *J. Neurophysiol.* 97, 26–43. doi:10.1152/jn.00720.2006
- Marquardt, T., McAlpine, D., 2006. A pi-limit for coding ITDs : implications for binaural models, in: Kollmeier, B., Klump, G., Hohmann, V., Langemann, U., Mauermann, M., Uppenkamp, S., and Verhey, J. (Ed.), *Hearing – from Basic Research to Applications*. Springer Berlin Heidelberg.
- Marquardt, T., McAlpine, D., 2009. Masking with interaurally “double-delayed” stimuli: The range of internal delays in the human brain. *J. Acoust. Soc. Am.* 126, EL177-EL182. doi:10.1121/1.3253689
- May, T., Van De Par, S., Kohlrausch, A., 2011. A probabilistic model for robust localization based on a binaural auditory front-end. *IEEE Trans. Audio, Speech Lang. Process.* 19, 1–13. doi:10.1109/TASL.2010.2042128
- McAlpine, D., Jiang, D., Palmer, A.R., 2001. A neural code for low-frequency sound localization in mammals. *Nat. Neurosci.* 4, 396–401.
- Meddis, R., 1986. Simulation of mechanical to neural transduction in the auditory receptor. *J Acoust Soc Am* 79, 702–711.
- Mi, J., Colburn, H.S., 2016. A Binaural Grouping Model for Predicting Speech Intelligibility in Multitalker Environments. *Trends Hear.* 20, 1–12. doi:10.1177/2331216516669919
- Middlebrooks, J.C., Clock, A.E., Xu, L., Green, D.M., 1994. A Panoramic Code for Sound Location by Cortical Neurons. *Science* (80-). 264, 842–844.
- Mountain, D., Anderson, D., Bresnahan, G., Deligeorges, S., Hubbard, A., Vajda, V., 2005. EarLab: large-scale simulation of auditory pathways, in: *Proc. Assoc. Res. Otolaryngol.*
- Osman, E., 1973. Correlation model of binaural detection: interaural amplitude ratio and phase variation for signal. *J Acoust Soc Am* 54, 386–389. doi:10.1121/1.1913589
- Pastore, M.T., Braasch, J., 2015. The precedence effect with increased lag level. *J. Acoust. Soc. Am.* 138, 2079–2089. doi:10.1121/1.4929940
- Patterson, R.D., Robinson, K., Holdsworth, J., McKeown, D., Zhang, C., Allerhand, M., 1992. Complex sounds and auditory images. *Simulation* 83, 429–446. doi:10.1016/B978-0-08-041847-6.50054-X
- Pecka, M., Brand, A., Behrend, O., Grothe, B., 2008. Interaural time difference processing in the mammalian medial superior olive: the role of glycinergic inhibition. *J. Neurosci.* 28, 6914–25. doi:10.1523/JNEUROSCI.1660-08.2008
- Raatgever, J., 1980. On the binaural processing of stimuli with different interaural phase relations. *Delft*.
- Rayleigh, L., 1907. On our perception of sound direction. *Philos Mag* 13, 214–232.
- Sandel, T.T., 1955. Localization of Sound from Single and Paired Sources. *J. Acoust. Soc. Am.* 27, 842. doi:10.1121/1.1908052
- Saremi, A., Beutelmann, R., Dietz, M., Ashida, G., Kretzberg, J., Verhulst, S., 2016. A comparative study of seven human cochlear filter models. *Jasa* 140, 1618–1634. doi:10.1121/1.4960486
- Sayers, B.M., Cherry, E.C., 1957. Mechanism of binaural fusion in the hearing of speech. *J Acoust Soc Am* 29, 973–987. doi:10.1121/1.1914990
- Schoenmaker, E., Brand, T., van de Par, S., 2016. The multiple contributions of interaural differences to improved speech intelligibility in multitalker scenarios. *J. Acoust. Soc. Am.* 139, 2589–2603. doi:10.1121/1.4948568
- Schröder, M.R., 1977. New viewpoints in binaural interactions., in: Evans, E. F. ; Wilson, J.P. (Ed.), *Psychophysics and Physiology of Hearing*. Academic Press, New York, New York, pp. 455–467.
- Seneff, S., 1988. A joint synchrony/mean-rate model of auditory processing. *J. Phon.* 16, 55–76.
- Shackleton, T.M., Meddis, R., Hewitt, M.J., 1992. Across frequency integration in a model of lateralization. *J. Acoust. Soc. Am.* 91, 2276. doi:10.1121/1.403663
- Shamma, S. a, Shen, N.M., Gopaldaswamy, P., 1989. Stereausis: binaural processing without neural delays. *J. Acoust. Soc. Am.* 86, 989–1006. doi:10.1121/1.398734
- Skottun, B.C., Shackleton, T.M., Arnott, R.H., Palmer, a R., 2001. The ability of inferior colliculus neurons to signal differences in

- interaural delay. *Proc. Natl. Acad. Sci. U. S. A.* 98, 14050–14054. doi:10.1073/pnas.241513998
- Slaney, M., 1998. Auditory toolbox, version 2.
- Slaney, M., 1988. Lyon's Cochlear Model.
- Søndergaard, P.L., Majdak, P., 2013. The Technology of Binaural Listening. doi:10.1007/978-3-642-37762-4
- Stange-Marten, A., Nabel, A.L., Sinclair, J.L., Fischl, M., Alexandrova, O., Wohlfrom, H., Kopp-scheinflug, C., Pecka, M., Grothe, B., 2017. Input timing for spatial processing is precisely tuned via constant synaptic delays and myelination patterns in the auditory brainstem. *Proc Natl Acad Sci USA* 1–8. doi:10.1073/pnas.1702290114
- Stecker, G.C., Bibee, J.M., 2014. Nonuniform temporal weighting of interaural time differences in 500 Hz tones. *J. Acoust. Soc. Am.* 135, 3541–7. doi:10.1121/1.4876179
- Stecker, G.C., Brown, A.D., 2010. Temporal weighting of binaural cues revealed by detection of dynamic interaural differences in high-rate Gabor click trains. *J. Acoust. Soc. Am.* 127, 3092–103. doi:10.1121/1.3377088
- Stecker, G.C., Ostreicher, J.D., Brown, A.D., 2013. Temporal weighting functions for interaural time and level differences. III. Temporal weighting for lateral position judgments. *J. Acoust. Soc. Am.* 134, 1242–52. doi:10.1121/1.4812857
- Stern, R. M., Jr. (1976). Lateralization, Discrimination, and Detection of Binaural Pure Tones, PhD thesis, Massachusetts Institute of Technology, December, 1976.
- Stern, R.M., Colburn, H.S., 1978. Theory of binaural interaction based on auditory-nerve data. IV. A model for subjective lateral position. *J Acoust Soc Am* 64, 127–140. doi:10.1121/1.381978
- Stern, R.M., Shear, G.D., 1996. Lateralization and detection of low-frequency binaural stimuli: Effects of distribution of internal delay. *J Acoust Soc Am* 100, 2278–2288. doi:10.1121/1.417937
- Stern, R.M., Trahiotis, C., 1997. Binaural mechanisms that emphasize consistent interaural timing information over frequency, in: Palmer, A.R., Rees, A., Summerfield, A.Q., Meddis, R. (Eds.), 11th International Symposium on Hearing: Auditory Physiology and Perception. Whurr Publishers, London, pp. 396–402.
- Stern, R.M., Trahiotis, C., 1992. The role of consistency of interaural timing over frequency in binaural lateralization, in: Cazals, Y.; Demany, L.; Horner, K. (Ed.), Auditory Physiology and Perception. Pergamon, New York, pp. 547–554.
- Stern, R.M., Zeiberg, A.S., Trahiotis, C., 1988. Lateralization of complex binaural stimuli: A weighted-image model. *J Acoust Soc Am* 84, 156–165. doi:10.1121/1.396982
- Sumner, C.J., O'Mard, L.P., Lopez-Poveda, E.A., Meddis, R., 2003. A Nonlinear Filter-Bank Model of the Guinea-Pig Cochlear Nerve: Rate Responses. *J. Acoust. Soc. Am.* 113, 3264–3274. doi:10.1121/1.1568946
- Takanen, M., Santala, O., Pulkki, V., 2014. Visualization of functional count-comparison-based binaural auditory model output. *Hear. Res.* 309, 147–63. doi:10.1016/j.heares.2013.10.004
- Teas, D.C., 1962. Lateralization of acoustic transients. *J. Acoust. Soc. Am.* 685, 1460–1465.
- Thompson, S.K., von Kriegstein, K., Deane-Pratt, A., Marquardt, T., Deichmann, R., Griffiths, T.D., McAlpine, D., 2006. Representation of interaural time delay in the human auditory midbrain. *Nat Neurosci* 9, 1096–1098. doi:10.1038/nn1755
- Trahiotis, C., Stern, R.M., 1989. Lateralization of bands of noise: Effects of bandwidth and differences of interaural time and phase. *J Acoust Soc Am* 86, 1285–1293. doi:10.1121/1.398743
- Trahiotis, C., Stern, R.M., 1994. Across-frequency interaction in lateralization of complex binaural stimuli. *J. Acoust. Soc. Am.* 96, 3804–3806. doi:10.1121/1.410570
- van Bergeijk, W.A., 1962. Variation on a Theme of Bekesy : A Model of Binaural Interaction. *J Acoust Soc Am* 34, 1431–1437.
- van der Heijden, M., Lorteije, J. a M., Plauška, A., Roberts, M.T., Golding, N.L., Borst, J.G.G., 2013. Directional hearing by linear summation of binaural inputs at the medial superior olive. *Neuron* 78, 936–48. doi:10.1016/j.neuron.2013.04.028
- van der Heijden, M., Trahiotis, C., 1999. Masking with interaurally delayed stimuli: the use of “internal” delays in binaural detection. *J. Acoust. Soc. Am.* 105, 388–99.
- Verhulst, S., Dau, T., Shera, C.A., 2012. Nonlinear time-domain cochlear model for transient stimulation and human otoacoustic emission. *J. Acoust. Soc. Am.* 132, 3842–3848. doi:10.1121/1.4763989
- von Békésy, G., 1930. Über das Richtungshören bei einer Zeitdifferenz oder Lautstärkenungleichheit der beiderseitigen Schalleinwirkungen. *Phys Z.* 31, 824–835.
- Wang, L., Colburn, H.S., 2012. A modeling study of the responses of the lateral superior olive to ipsilateral sinusoidally amplitude-modulated tones. *J. Assoc. Res. Otolaryngol.* 13, 249–67. doi:10.1007/s10162-011-0300-5
- Wang, L., Devore, S., Delgutte, B., Colburn, H.S., 2014. Dual sensitivity of inferior colliculus neurons to ITD in the envelopes of high-frequency sounds: experimental and modeling study. *J. Neurophysiol.* 111, 164–81. doi:10.1152/jn.00450.2013
- Wierstorf, H., Raake, A., Spors, S., 2013. The Technology of Binaural Listening. doi:10.1007/978-3-642-37762-4
- Yao, J.D., Bremen, P., Middlebrooks, J.C., 2015. Transformation of Spatial Sensitivity along the Ascending Auditory Pathway. *J. Neurophysiol.* 35, jn.01029.2014. doi:10.1152/jn.01029.2014
- Yin, T.C., Chan, J.C., 1990. Interaural time sensitivity in medial superior olive of cat. *J. Neurophysiol.* 64, 465–88.
- Yost, W.A., 1970. Tone-on-tone Binaural Masking. Indiana University.
- Zhang, X., Heinz, M.G., Bruce, I.C., Carney, L.H., 2001. A phenomenological model for the responses of auditory-nerve fibers: I. Nonlinear tuning with compression and suppression. *J. Acoust. Soc. Am.* 109, 648. doi:10.1121/1.1336503

- Zhou, Y., Carney, L.H., Colburn, H.S., 2005. A model for interaural time difference sensitivity in the medial superior olive: interaction of excitatory and inhibitory synaptic inputs, channel dynamics, and cellular morphology. *J. Neurosci.* 25, 3046–58. doi:10.1523/JNEUROSCI.3064-04.2005
- Zilany, M.S.A., Bruce, I.C., Carney, L.H., 2014. Updated parameters and expanded simulation options for a model of the auditory periphery. *J. Acoust. Soc. Am.* 135, 283–286. doi:10.1121/1.4837815
- Zilany, M.S.A., Bruce, I.C., Nelson, P.C., Carney, L.H., 2009. A phenomenological model of the synapse between the inner hair cell and auditory nerve: Long-term adaptation with power-law dynamics. *J. Acoust. Soc. Am.* 126, 2390–2412. doi:10.1121/1.3238250
- Zwislocki, J., Feldman, R.S., 1956. Just Noticeable Differences in Dichotic Phase. *J Acoust Soc Am* 28, 860–864.