bioRxiv preprint doi: https://doi.org/10.1101/687178; this version posted September 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

General neural mechanisms can account for rising slope preference in localization of ambiguous sounds

Jean-Hugues Lestang[1,*] and Dan F. M. Goodman[1]

[1] *Department of Electrical and Electronic Engineering, Imperial College London*

Sound localization in reverberant environments is a difficult task that human listeners perform effortlessly. Many neural mechanisms have been proposed to account for this behavior. Generally they rely on emphasizing localization information at the onset of the incoming sound while discarding localization cues that arrive later. We modelled several of these mechanisms using neural circuits commonly found in the brain and tested their performance in the context of experiments showing that, in the dominant frequency region for sound localisation, we have a preference for auditory cues arriving during the rising slope of the sound energy (Dietz et al., 2013). We found that both single cell mechanisms (onset and adaptation) and population mechanisms (lateral inhibition) were easily able to reproduce the results across a very wide range of parameter settings. This suggests that sound localization in reverberant environments may not require specialised mechanisms specific to perform that task, but could instead rely on common neural circuits in the brain. This would allow for the possibility of individual differences in learnt strategies or neuronal parameters. This research is fully reproducible, and we made our code available to edit and run online via interactive live notebooks.

Introduction

The precedence effect, or law of the first wave, is a psychoacoustical phenomenon occurring when multiple sounds reach the listener's ears in quick succession (Wallach et al., 1949). For small enough delays, the perceived sound is located near the source of the leading sound. In other words, in order to extract reliable localization cues, the auditory system favors the computation of dichotic cues conveyed in the first sound arriving at the ears. Doing so unburdens the localization process from dealing with confounding reverberations. Although well studied (Zurek, 1987; Blauert, 1997; Litovsky et al., 1999; Brown et al., 2015), the neural mechanisms behind the precedence effect are still not well understood. In particular, it is not clear at what stage of the auditory pathway the re-weighting of the interaural cues occurs.

Principally, two main processing sites were suggested. On the one hand, different types of neural adaptation have been reported to take place in the periphery of the auditory pathway. Proposed models of peripheral adaptation range from simulating the ringing of the basilar membrane (Tollin, 1998) to adaptation occurring at the synapse between the inner hair cells and the auditory nerve fibers (Hartung and Trahiotis, 2001; Xia and Shinn-Cunningham, 2011). Similarly, onset cells located in the cochlear nucleus (Rothman and Manis, 2003; Spencer et al., 2012, 2018) are likely to play a role in discarding counfounding localization cues. On the other hand, models of the auditory system including lateral inhibitory circuits in the medial superior olive (MSO) or inferior colliculus (IC) were also able to account for some aspects of the precedence effect. Most of these models rely on the transient inhibition of the lagging sound (Lindemann, 1986a,b; Zurek, 1987; Xia et al., 2010). Overall, peripheral models seem to account well for precedence effect related behaviors with short sounds while inhibition based models produce better results with longer sounds (Braasch and Blauert, 2003).

To further understand how interaural time differences (ITDs) are processed by the auditory system in reverberant environments, a study from Dietz et al. (2013), showed that human listeners tend to rely on interaural phase differences (IPDs) located in the onset of low-frequency amplitude modulated sounds. To demonstrate this, the authors crafted an amplitude modulated binaural beats (AMBB) stimulus exhibiting a time-varying IPD performing a full cycle (from 0° to 360°) over

bioRxiv preprint doi: https://doi.org/10.1101/687178; this version posted September 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

the time-course of a single envelope cycle, while ensuring that there is no interaural level difference. By asking participants to match the perceived location produced by the AMBB sounds to static IPD pointers, Dietz et al. (2013) observed that the subjects mostly made use of the interaural cues positioned in the rising portion of the envelope of the stimulus (instead of the peak of the envelope where the amplitude is maximal). The perceived IPD was also shown to increase with modulation frequency. Later on, Hu et al. (2017) conducted a similar experiment and uncovered that IPD extraction in the rising slope is conditional to the value of the carrier frequency. For 200 Hz stimuli, the subjects mostly reported IPDs located at the peak of the envelope. However, for 600 Hz stimuli, subjects prioritized the rising portion of the envelope.

We investigated a variety of neural mechanisms that could account for the perceived location of the AMBB stimulus in an attempt to rule out certain mechanisms. Instead, we found that all the mechanisms we tried were able to account for this effect. This raises the possibility that the effect may be the result of general processes occurring throughout the brain rather than specialised mechanisms for sound localisation in reverberant environments, and would allow for considerable individual differences in terms of neuronal parameters (Prinz et al., 2004) or localization strategies (Keating et al., 2016). In the first half of this paper, we investigate single neuron mechanisms that occur before binaural integration takes place, using a simple, general model including multiple mechanisms that can enhance onset (including adaptation and delayed inhibition). In the second half of the paper we investigate population level mechanisms based on the interaction of excitation and inhibition after binaural integration. To ensure computational reproducibility, the entirety of the code used to produce the results presented in this paper is available online (`https://github.com/neural-reckoning/simple_ambb_modelling`) and can be edited and run directly in the browser.

Results

We investigated the neural mechanisms that could account for the data collected by Mathias Dietz and colleagues on amplitude modulated binaural beats (AMBBs; Dietz et al. 2013, 2014). These are sinusoidally amplitude modulated tones where there is an interaural phase difference (IPD) in the tone or carrier that linearly increases from 0 to 360° during each amplitude modulation cycle. At low modulation frequencies this is perceived as a sound moving around the head. In psychophysical experiments, subjects were asked to move a slider to modify a second amplitude modulated tone with a static IPD (controlled by the slider) to match as closely as possible the AMBB. It might be expected that when forced to assign a single IPD to a stimulus that has all possible IPDs, the IPD at the time where the stimulus was strongest would be chosen (corresponding to an IPD of 180°). However, they found that subjects tended to choose an IPD in the rising portion of the envelope (figure 1), at a phase that increases monotonically with the modulation frequency but is always less than 180°.

We consider two possible explanations for this effect based on mechanisms that occur either before or after binaural integration. In Single neuron mechanisms, we investigate whether monaural single neuron properties could explain the observations, and what conditions the observed data would impose on the parameters and types of those neurons. Dietz et al. (2014) argue that the mechanism must occur before binaural integration takes place, but their argument is based on only a single neuron. With a population of neurons, an inhibitory mechanism can also explain the effect. In Population mechanisms, we investigate these population level effects with a lateral inhibition mechanism. Throughout, we use abstract rate models of neurons with rich dynamics to extract the essential details while keeping the model complexity manageable (similarly to Goodman et al. 2017). This enables us to investigate how the behaviour of the model depends on all of the parameters and plot the parameter spaces, which would not be possible with a biophysically detailed model with a large number of parameters.

Single neuron mechanisms

A natural place to start given that subjects tend to favour the rising slope would be to assume that the input signal is being differentiated, and therefore the reported IPD will be at the peak of the (differentiated) neural signal (figure 1A). However, this peak occurs at 90° regardless of the

bioRxiv preprint doi: https://doi.org/10.1101/687178; this version posted September 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

modulation frequency, which does not match the pattern of a later preference at higher modulation frequencies observed experimentally, and also leaves open the question of the neural mechanism implementing this differentiation.

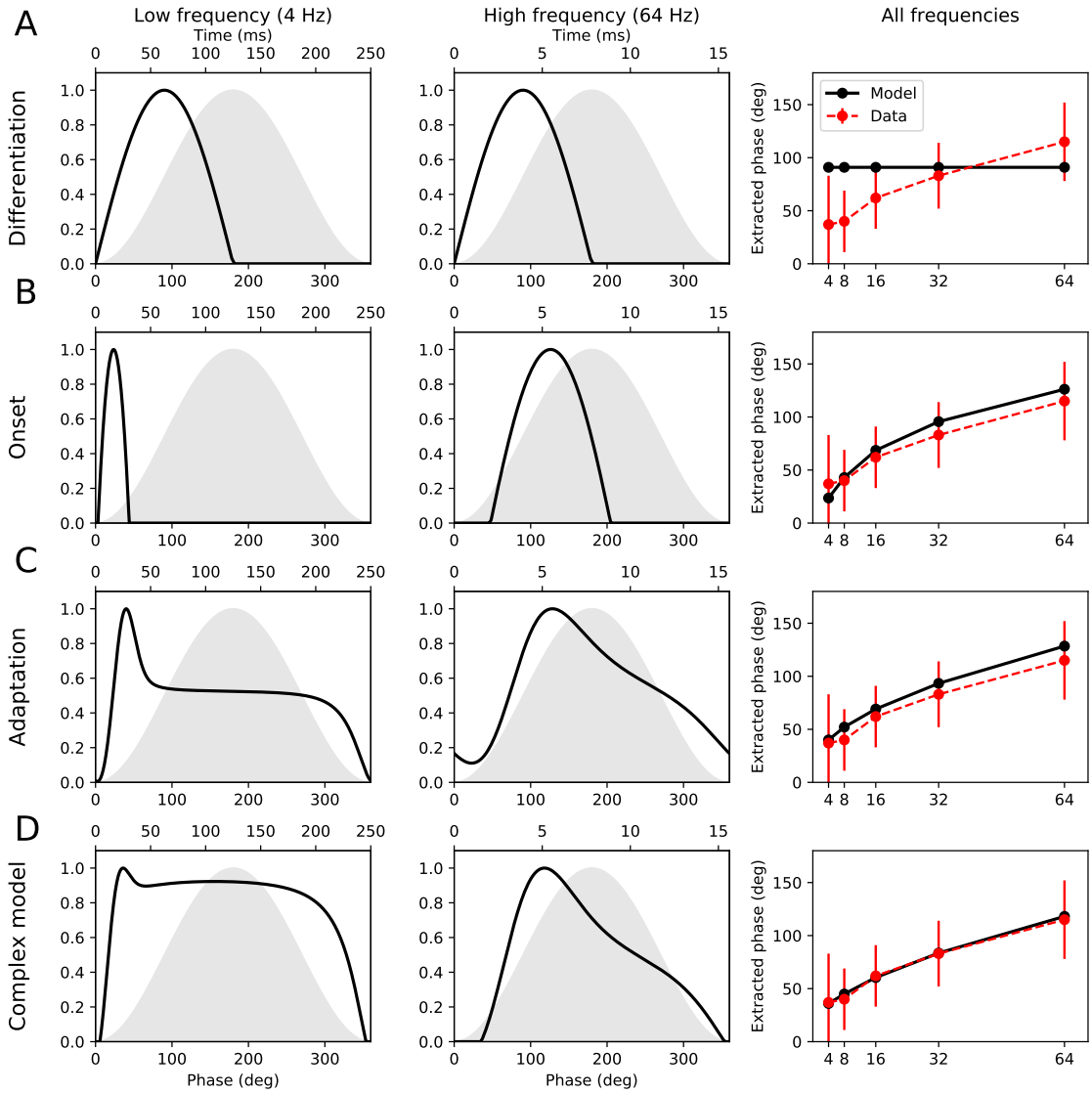

Figure 1: Basic mechanisms of the single neuron model. In each panel, the left and middle plots show, respectively, the response of the model at a low (4 Hz) and high (64 Hz) modulation frequency. The envelope is shown as the grey shaded area, while the output of the model is shown as a black line. The right hand plot shows the location of the peak response of the model at different modulation frequencies (black line), and the data for human subjects (red dashed line). The human data shows the standard deviation of responses across subjects as error bars. (A) Model output is the rectified differential of the envelope. (B) Onset model. (C) Adaptation model. (D) Complex model using both onset and adaptation mechanisms.

Onset and adaptation mechanisms can explain rising slope preference

We model onset behaviour abstractly as the weighted difference of two low-pass filtered signals (with different time constants). This can be interpreted either as two precisely timed excitatory/inhibitory pathways, or in a similar way to the octopus cell model of Spencer et al. (2012, 2018) and Ferragamo and Oertel (2002). Their model uses a spike threshold on the rate of change of the membrane potential (Platkiewicz and Brette, 2010, 2011), which we reformulated in terms of firing rates (see Methods and Appendix). With this model it is possible to closely fit the experimentally observed relationship between the modulation frequency and the extracted phase with this model (figure 1B). The parameters are $\beta \geq 0$ the relative weight of the subtracted signal (so $\beta = 0$ corresponds

bioRxiv preprint doi: https://doi.org/10.1101/687178; this version posted September 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

to just low pass filtering the signal), and the time constants of the positive and negative filters (τ_e and τ_i respectively).

Another set of mechanisms relate to various forms of adaptation, e.g. spike frequency adaptation or synaptic depression. We model these with a single reservoir model that exhibits a response to a constant input that exponentially decays from an initial to a fully adapted response. This is similar in structure to several well used and equivalent models of adaptation in the auditory nerve (Meddis, 1986; Westerman and Smith, 1988; Zhang and Carney, 2005). It is also equivalent to the short-term synaptic plasticity model of Tsodyks et al. (1998) (with depression only, no facilitation, for the derivation see appendix A of Tsodyks and Wu 2013). Our adaptation model is therefore able to summarise a number of adaptation-related mechanisms, and uses only two parameters (see Methods). Figure 1C shows that this adaptation mechanism is also able to closely fit the data. The parameters are τ_a the time constant of the adaptation, and $0 \leq \alpha < 1$ the strength of the adaptation (with $\alpha = 0$ indicating no adaptation).

In our full model, both onset and adaptation mechanisms are available, as well as some additional mechanisms such as gain and compression, and this enables us to fit the data even more closely (figure 1D).

Mechanisms are highly robust

We computed which values of the model parameters were consistent with the data, and found that the model was very robust, with large regions of the parameter space giving good fits (Figure 2A). Note that since the model has 6 free parameters and there are only 5 data points, it is a priori unsurprising that it is possible to fit the data well. However, this is not a case of simple curve fitting as the model is highly constrained in terms of the possible phase/f_m curves it can produce (Figure 2B), with the majority (73%) being monotonically increasing, for example. We investigate this in more detail in Curve fitting analysis below. Finally, if we remove some mechanisms and their associated parameters, in almost all cases the model is able to fit the data well with as few as three parameters (Figure 3), and a mathematical approximation fits reasonably well with just one parameter (see Appendix).

Using this model, we can determine which mechanisms are able to explain the data, and which mechanisms are essential. We considered model variants where some of the parameters were given fixed values (Figure 2A) or some mechanisms removed (Figure 3). The key mechanisms of the model are adaptation and onset: either mechanism can work, although the best fits obtained with adaptation only are better than the best fits obtained with the onset mechanism only (Figure 3ABC); and at least one of the two mechanisms must be present (Figure 2A, α versus β). The compression and gain mechanisms did not contribute substantially (Figure 3A). We investigated the octopus cell model discussed earlier (Ferragamo and Oertel, 2002; Spencer et al., 2012, 2018) and found that it could provide a good fit only if adaptation was present alongside the onset mechanism (Figure 3AD). For model variants including the onset mechanism, the positive or excitatory time constant τ_e must be small (Figures 2A and 4B), and the negative or inhibitory time constant τ_i must be larger than τ_e (Figure 3CD, τ_e versus τ_i). An interesting feature of all the model variants is that the extracted phase curve in the experimental data appears to be at or close to the lowest values possible in the models (Figure 3A).

Best model fits divide into two main phenomenological cell types

We investigated the parameters and phenomenological behaviour of all model cells that were capable of reproducing the experimental data within an error of 30° (Figure 4), as this was the size of the smallest error bar in the experimental data. We computed the rate and temporal modulation transfer functions (rMTF, tMTF), and extracted the corresponding best modulation frequencies (BMF), mean MTFs, and modulation depths (MD) for each cell. We found that there was no clear demarcation of the parameter space into discrete cell types based on parameter values, as a continuum of values led to good fits (Figure 4A). However, we found that there was a clear demarcation into two cell types based on their modulation transfer functions. One cell type has a high tMTF: it phase locks sharply to the stimulus envelope at all modulation frequencies (high mean tMTF value, low tMD), and has a strongly modulated high pass or band-pass rMTF. This

bioRxiv preprint doi: https://doi.org/10.1101/687178; this version posted September 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

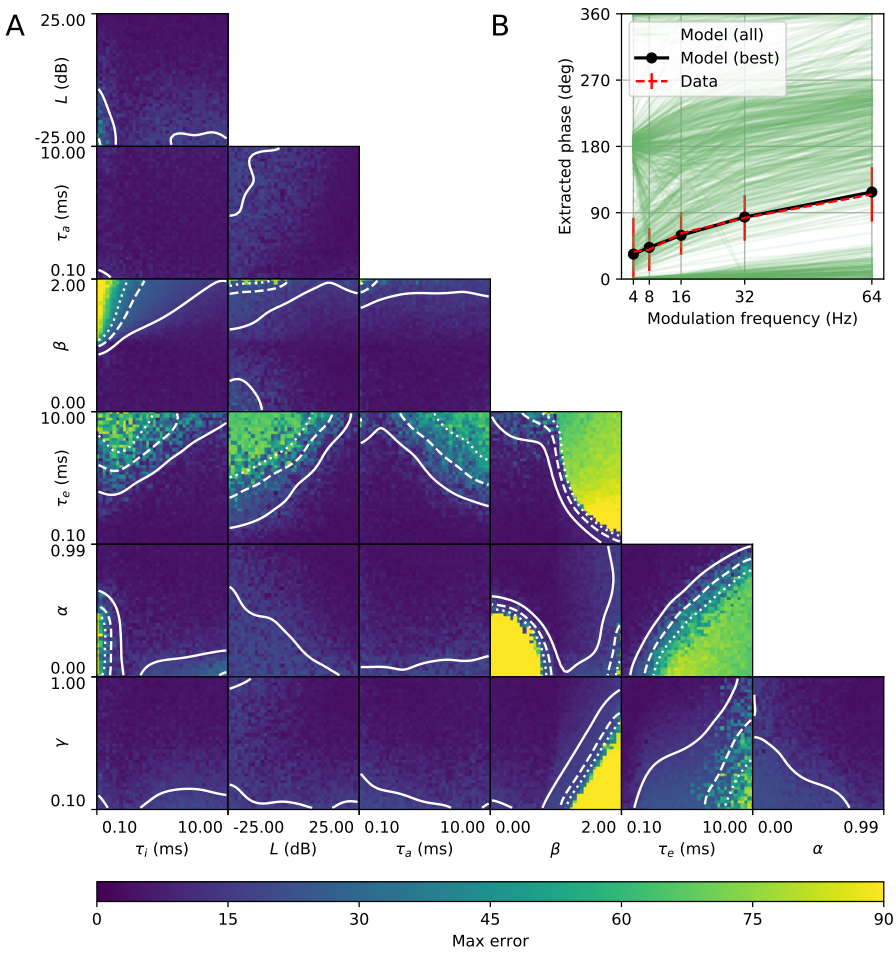

Figure 2: Parameter space. (A) Model error plotted as a function of different pairs of parameters. Each pixel shows the lowest error achievable for a particular fixed value for the corresponding pair of parameters. White contours show 15 degree error (solid), 30 degrees (dashed) and 45 degrees (dotted). (B) A sample of the extracted phase curves for 1000 randomly selected parameters (green curves, excluding parameters which give zero output for some value of f_m), the data (red dashed) and the best fitting model (black).

cell type corresponds to the parameter $\beta > 1$, or a strong onset mechanism which allows it to strongly phase lock to the envelope. These cells can have a wide range of adaptation strengths, fast excitatory time constants and slower inhibitory time constants. The other cell type has a more variable tMTF with a correspondingly higher tMD, along with a typically lower rMD. This cell type corresponds to the parameter $\beta < 1$, or a weak onset mechanism that is consistent with a weaker locking to the envelope. These cells tend to have stronger adaptation, and are consistent with a wider range of excitatory and inhibitory time constants compared to the onset cell type, as well as a wider range of behaviours (Figure 4B). Both cell types were about equally present in the parameter space considered.

Homogeneous neurons without onset mechanisms do not exhibit an early preference

So far, we have considered a model that only responds to the envelope of the sound, which can be seen either as an approximation, or as applying only to cells with low synchronisation to the carrier. We added the carrier to the model, and one additional low pass filter mechanism (representing the inner hair cell or other low pass filtering processes). For simplicity we used a low pass filter with time constant τ_{IHC}, although higher order filters may provide a better fit to the data (Russell and Sellick, 1983). In addition to considering the carrier frequency of $f_c = 500$ Hz used in Dietz et al. (2013, 2014), we also calculated results for $f_c = 200$ Hz. This lower carrier frequency was studied in Hu et al. (2017) with a similar experimental design, and they found that there didn't appear to be an extraction of the early IPD at this f_c, only at the higher f_c (they used 600 Hz in that

bioRxiv preprint doi: https://doi.org/10.1101/687178; this version posted September 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

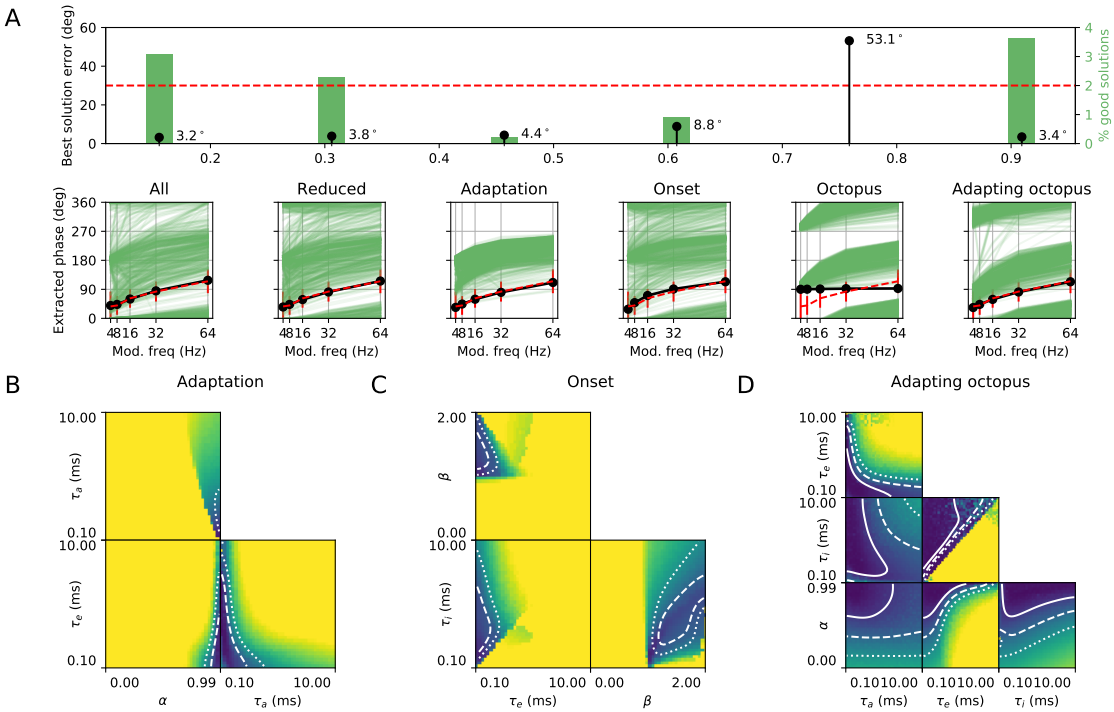

Figure 3: Comparison of restricted models. (A) Extracted phase curves (lower panels as in Figure 2B) for different model variants: *All* parameters available; a *Reduced* set where gain and compression were removed ($L = 0$, $\gamma = 1$), which is also the case in the remaining variants; *Adaptation* only ($\beta = 0$); *Onset* only ($\alpha = 0$); *Octopus* cell ($\alpha = 0$, $\beta = 1$); *Adapting octopus* cell ($\beta = 1$). The upper panel shows the error of the best performing set of parameters found for that variant. All are within the error bars of the data (dashed red line) except for the octopus cell. (B-D) Parameter maps as in Figure 2A for three of the model variants (adaptation, onset, adapting octopus).

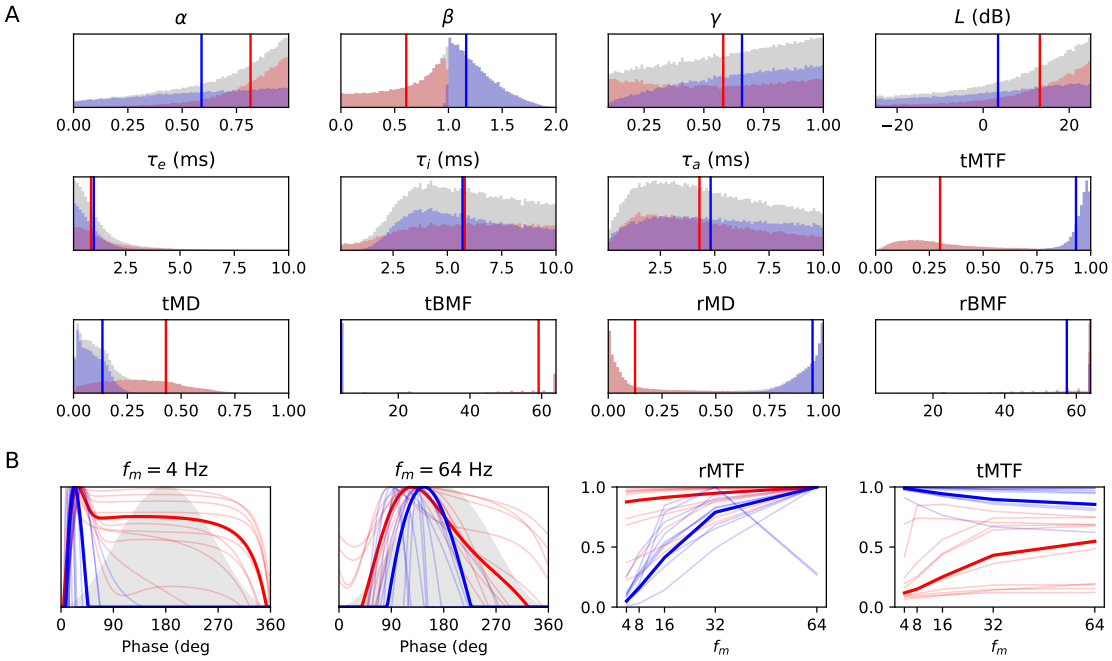

Figure 4: Cell types. (A) Histograms for all parameters and properties leading to an error of less than 30° (grey), for the strong onset cell class with tMTF$>$ 0.75 (blue), or the weak onset cell class with tMTF$<$ 0.75 (red). Vertical lines show mean values for the two classes. (B) Sample responses from the two cell classes. Left two panels as in Figure 1, right two panels show detailed rate and temporal modulation transfer functions. Thin lines show a random sample of curves from each class, and thick lines show the most representative example from that class (smallest distance to all other points in the cell class parameter space).

bioRxiv preprint doi: https://doi.org/10.1101/687178; this version posted September 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

paper, similar to the 500 Hz used in the earlier papers). However, the results are not directly comparable, so we decided to measure the error in the model as the combination of the error at 500 Hz as measured above, and an error at 200 Hz assuming that the equivalent experimental results were a flat IPD = 180°. The contribution of the error at 200 Hz was weighted less as there is no direct experimental data in this case, only a hypothetical curve based on a similar experiment (see Methods for more details).

We found that the model was able to simultaneously reproduce the two different extracted phase curves at $f_c = 200, 500$ Hz (Figure 5). The main difference in the parameters found is that with the carrier frequency included in the model it is not possible to closely fit the data with an adaptation-only model ($\beta = 0$), and indeed onset strength has to be quite high ($\beta > 1$, Figure 5B). In addition, some well fitting parameters in the envelope only model with a large τ_i no longer work well in this case. We found that uniformly the tMTF was higher at $f_c = 500$ Hz compared to 200 Hz, the tMD lower, and the rMD was always very high for $f_c = 500$ Hz (but could have a wide range of values at 200 Hz).

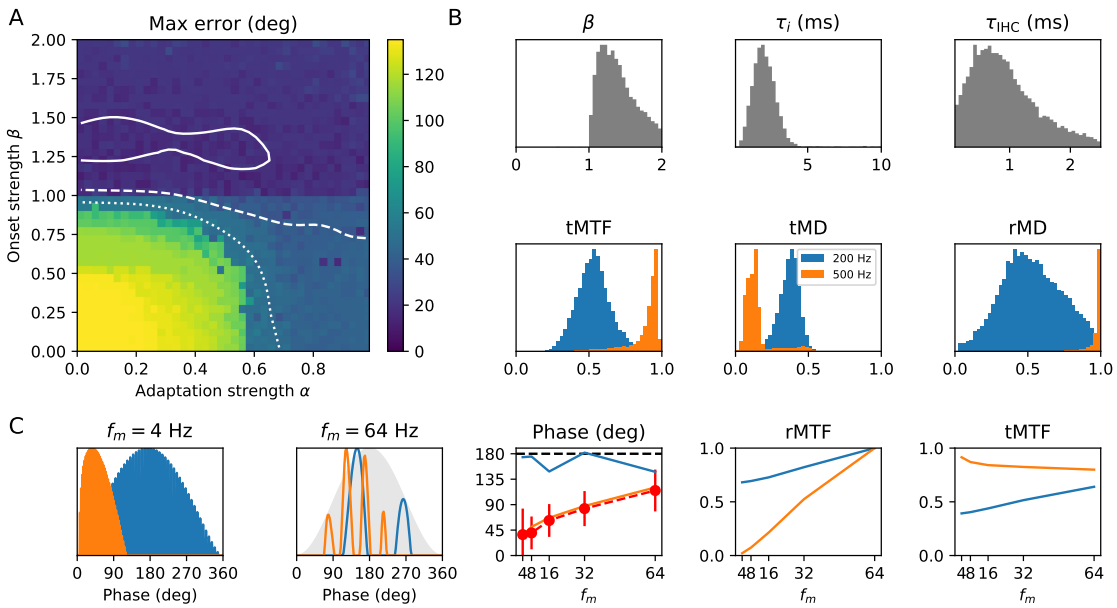

Figure 5: Effect of introducing carrier frequency. (A) Error in best fit for fixed adaptation strength α and inhibition strength β (as in Figure 4). Here, the error is measured at both $f_c = 500$ Hz and $f_c = 200$ Hz assuming a hypothetical flat 180° extracted in the latter case. The 200 Hz error is weighted less as there is no corresponding experimental data. (B) Parameter and property histograms as in Figure 4. Here, colours show properties at 200 Hz (blue) and 500 Hz (orange). (C) Best fit in detail as in figure 4.

It would be tempting to conclude from this that only onset cells can account for all the experimental data of Dietz et al. (2013, 2014) and Hu et al. (2017), however the populations of cells responding to a 200 Hz and 500 Hz carrier frequency are different, and there is no a priori reason to think that they should be the same cell types or have the same parameters.

Model predicts earlier preference at higher sound levels

We took the collection of good parameters (with an error less than 30°) and applied different gains to amplify or attenuate the signal before recomputing the extracted phase curves. We found that with a considerable degree of consistency, increasing the sound level led to an earlier phase being extracted, while decreasing the level led to a later phase (Figure 6). Experimental data has not yet been collected to quantitatively test this hypothesis, but early observations are in qualitative agreement (Mathias Dietz, personal communication of unpublished observations).

bioRxiv preprint doi: https://doi.org/10.1101/687178; this version posted September 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

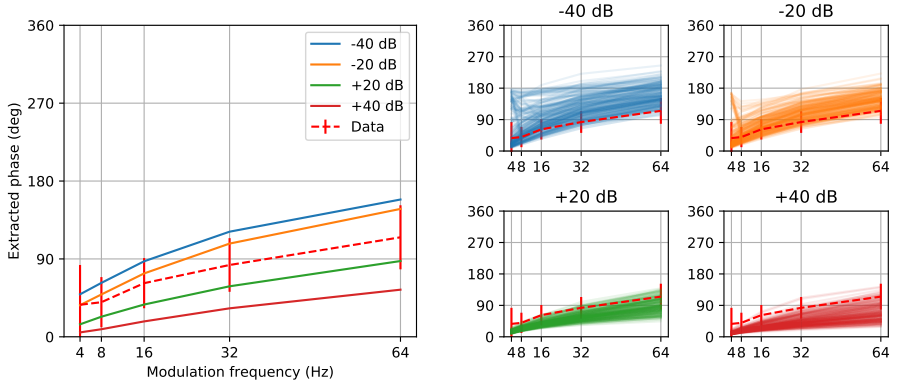

Figure 6: Single neuron model predictions for varying sound level. Left panel shows mean prediction at different sound levels for parameters that match the experimental data well at a sound level of +0 dB. Right panels show sample extracted phase curves for each sound level.

Population mechanisms

In order to investigate whether mechanisms acting after binaural integration could give rise to a similar effect, we next investigated the role of population mechanisms assuming no monaural mechanisms that give rise to an early phase preference. In the experiments of Dietz et al. (2013) participants were asked to choose which amplitude modulated tone with a static IPD sounded most similar to the amplitude modulated binaural beat (AMBB) stimulus. We therefore modelled the process as pattern matching on the neural population (as in the sound localisation model of Goodman et al. 2013). We included a lateral inhibition mechanism that is found throughout the brain, and in the case of sounds with a static IPD serves to better separate the patterns. In the case of the AMBB stimulus, this model was able to robustly reproduce the preference for early phase, getting later with increasing modulation frequency, and fit the experimental data with low error across a wide range of parameters (figure 7).

In more detail, the model receives as input responses from a population of cells with different IPD tuning, with the density of best IPDs as measured for guinea pigs (McAlpine et al., 2001). These neurons directly excite a second layer of neurons and inhibit the neighbours of those neurons, serving to sharpen the pattern of responses. Finally, we compare the patterns of activities for the AMBB stimulus with the patterns obtained for different amplitude modulated tones with static IPDs, and select the static IPD with the most similar response. The model has five parameters: τ_b is the time constant of the inhibition process; σ_k is the width of the lateral inhibition as a fraction of the whole population (so $\sigma_k = 1$ means all neurons inhibit all others and $\sigma_k = 0$ means no inhibition); γ is the strength of the inhibition; k is an integer that determines the shape of the IPD tuning curve and can be experimentally fit to different animal models; and m is the envelope synchronization of the inputs, so when $m = 0$ the inputs are not synchronised to the envelope at all, at $m = 1$ they exactly match the shape of the envelope, and for $m > 1$ they have enhanced synchronisation (e.g. onset cells). See Methods for further details of the model.

Lateral inhibition and pattern matching gives rise to a robust early phase preference

We calculated model fits for a wide range of parameter values (89,100 different parameter sets in total). The best fit had an error of just 5° (figure 7B), and large regions of the parameter space led to low errors (figure 7A) of less than 30° (the size of the smallest error bar in the experimental data). Overall, 8.1% of the parameter sets led to an error below 30° and 41% of the parameter sets led to solutions which increased monotonically with the modulation frequency. Most of the parameter sets (78%) led to a later phase preference than the experimental data, 2% to an earlier phase preference and 20% were earlier for some modulation frequencies and later for others.

Some parameters are more tightly constrained than others if we seek a close fit to the data. The synchronization index m for parameter sets with low error shows a particularly striking pattern (figure 7C). There is a very steep drop in the number of good solutions where $m < 1$, suggesting

bioRxiv preprint doi: https://doi.org/10.1101/687178; this version posted September 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

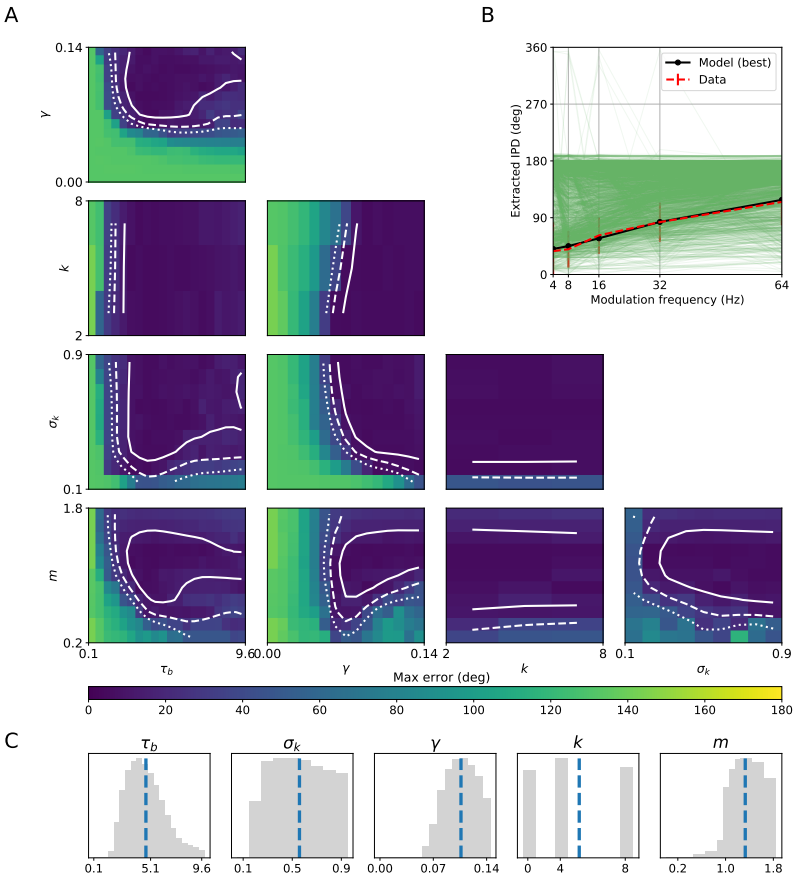

Figure 7: Population model parameter space, as in figure 2. (A) Model error plotted as a function of different pairs of parameters. Each pixel shows the lowest error achievable for a particular fixed value for the corresponding pair of parameters. White contours show 15° error (solid), 30° (dashed) and 45° (dotted). (B) Sample and best model results. (C) Histograms of the parameters leading to an error less than 30°.

that enhanced synchronisation may play a role. This enhanced synchronisation is evident in some auditory neurons, and may come about as a result of the onset mechanisms discussed in Single neuron mechanisms for example. For parameters τ_b and γ the best fits were obtained in a limited although not extreme range. Best fits were obtained for τ_b in the moderate range 2-8 ms, for example. Larger values of the inhibition strength γ lead to a shut down of the network and less inhibition leads to solutions closer to 180°. The least critical parameters were σ_k and k. The former is the width of the lateral inhibition, which could take on almost any value except those close to the extremes where there was no inhibition ($\sigma_k = 0$) or maximum inhibition ($\sigma_k = 1$). The tuning curve width, controlled by k, had almost no effect.

Distribution of tuning leads to peak preference at low carrier frequencies

Hu et al. (2017) suggests that carrier frequency has an important effect on the perceived IPD of the AMBB stimulus. They showed that with a carrier frequency of 200 Hz, the perceived IPD is shifted toward the peak of the envelope, while at 600 Hz it is perceived during the rising slope as in Dietz et al. (2013). McAlpine et al. (2001) found that the distribution of best IPDs at 200 Hz is different to that at 500 Hz. In our model, we found that for some parameters, this change in BIPD distribution alone caused a shift in perceived IPD towards the peak (figure 8). As before, when investigating model fits at 200 Hz and 500 Hz we weight the error at 200 Hz much less (one third) than the error at 500 Hz, as the experimental data from Hu et al. (2017) is much less precise (preferred IPD is only reported as rising, peak or falling, a granularity of about 90°). The lowest error found was approximately 15°, meaning that the extracted phase was within 15° for all modulation frequencies at 500 Hz, and within 45° at 200 Hz, both well within the limits set by the data. In order to interpret the sensitivity of the model to the BIPD distribution, we computed the

bioRxiv preprint doi: https://doi.org/10.1101/687178; this version posted September 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

mean preferred IPDs across all good solutions (error below 30°) for carrier frequencies from 200 to 1000 Hz. We found that decreasing the carrier frequency led to a shift in preference towards the peak (figure 8D). For carrier frequencies above 500 Hz, the preference did not shift any earlier.

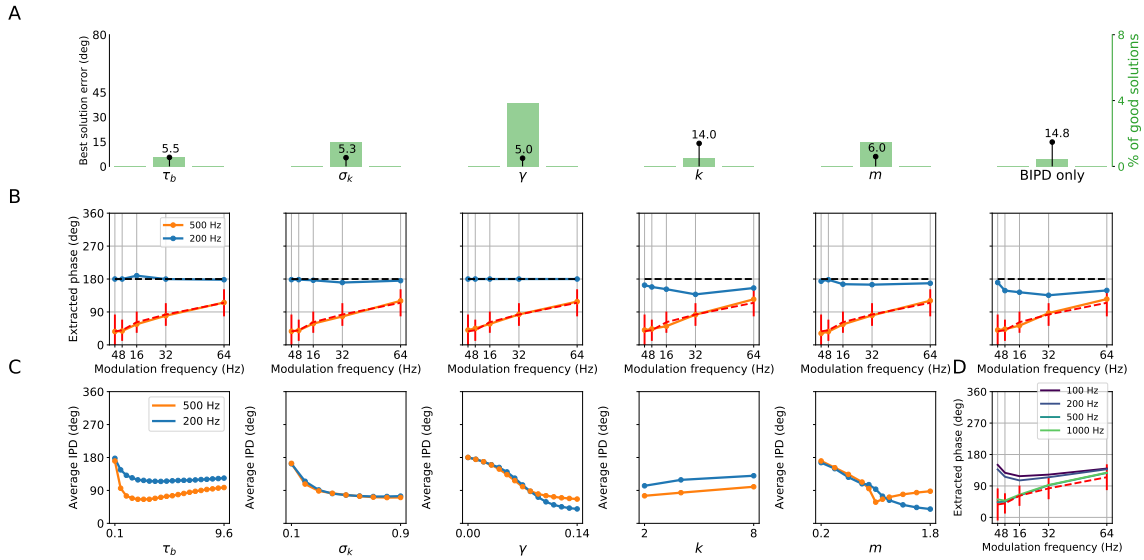

Figure 8: (A) Error in best solution (black lines and dots), and percent of the parameter space leading to good solutions (green bars), when allowing one parameter to take different values at 200 Hz and 500 Hz (except for "BIPD only" where all parameters are identical except for the BIPD distribution). (B) Best solutions found at 200 Hz and 500 Hz. (C) Trend in mean preferred IPD over modulation frequencies when varying the value of a single parameter. The trend was obtained by averaging over all solutions below 30°. (D) Mean IPD across all good solutions for various carrier frequencies.

Although the error for the best solution is low, this result requires some precise parameter tuning, and only 0.5% of parameters tested had an error lower than 30°. This model assumes that the parameters of the neurons must be the same at a carrier frequency of 200 and 500 Hz, but these are different neurons (due to tonotopy, Humphries et al. 2010; Ress and Chandrasekaran 2013), and they may therefore have different parameters. We refer to the case where all parameters must be the same as homogeneous and the case where they may be different as heterogeneous. We investigated heterogeneous networks by allowing only one parameter to have a different value at 200 and 500 Hz (figure 8). Unsurprisingly, this leads to much better fits. Allowing for different relative strengths of excitation and inhibition γ at 200 and 500 Hz – which is very plausible as synaptic weights can be learned – leads to a best fit with an error of just 5°, and almost 4% of the parameter space with errors less than 30°. Allowing σ_k and m to vary led to only slight larger minimum errors, although in these cases a somewhat smaller percentage of parameters led to errors below 30°. Allowing the tuning curve width to vary (controlled by k) did not contribute to a good fit. We investigated how each parameter contributes to the mean preferred IPD (calculated across all good solutions and all modulation frequencies, figure 8C). Preferred phase decreased monotonically when either the width (σ_k) or strength of inhibition (γ) increased, and tended to decrease with increasing synchronisation (m), except at 500 Hz for larger values of m. Very small time constants of inhibition (τ_b) led to a later phase preference, but had relatively little effect as long as the time constant was larger than around 1 ms. Narrower tuning curves (larger values of k) led to slightly later phase preference.

Curve fitting analysis

For both the single neuron and population models, we have around the same number of parameters as data points. We therefore investigated whether or not our results could be explained simply as curve fitting by calculating how well the model could fit the data if the data were different (figure 9). Two different conditions were used. In the first one, the target IPDs for each modulation frequency were chosen independently and uniformly at random between 0° and 360°. In the second condition, the target IPDs were again chosen randomly, but this time between 0° and 180°, and with the additional constraint that they should be monotonically increasing with modulation frequency. In both cases, the best fits obtained to the actual data were much better than the best fits for

bioRxiv preprint doi: https://doi.org/10.1101/687178; this version posted September 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

the random target IPDs in almost all cases. None of the 1,000 fully random IPD targets could be fitted as well as the experimental data. With the extra constraints, the single neuron model could fit the random data better than the experimental data only 3% of the time, and the population model only 8% of the time. Since the models were not able to fit the majority of random data, even when constrained to have a qualitatively similar shape to the experimental data, these results cannot simply be explained as curve fitting.

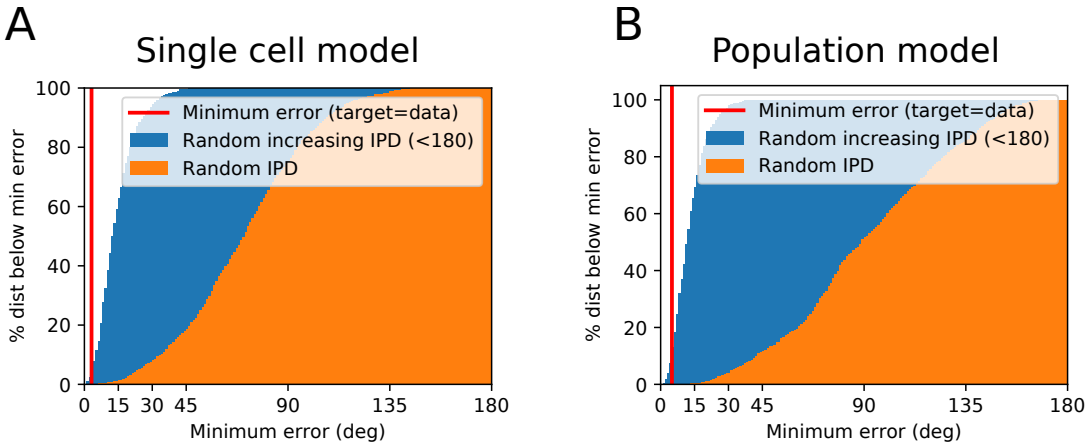

Figure 9: Best fits for experimental data (red line) and randomised target data (orange for fully random, blue when constrained to be monotonically increasing with modulation frequency less than 180°). Shaded areas show cumulative histograms of the minimum error found for the target data (so 0% of randomly selected data could be fit with an error of 0° while 100% could be fit with an error less than 180°). (A) Single cell model. (B) Population model.

Discussion

We modelled a number of single neuron and population level neural mechanisms across different stages of the auditory pathway to account for the preference for extracting binaural cues during the rising slope of an amplitude modulated signal (Dietz et al., 2013). By emphasizing the cues contained in the onsets of sounds, this effect was suggested to be important in responding to the sound following a direct path in a reverberant environment, and discarding later arriving information from indirect paths. This may be an important factor underlying a number of phenomena, including spatial release from masking in cocktail party situations (Cherry, 1953; Freyman et al., 1999).

We found that adaptation and onset mechanisms occurring in single neurons before binaural integration (for example in the auditory nerve or cochlear nucleus) could lead to a rising slope preference that was an excellent quantitative fit to the experimental data recorded by Dietz et al. (2013) across a very wide range of parameters. The best results fell into two main separate classes: strong onset or strong adaptation. Of these two, onset cells were more consistent with the result that preference shifts towards the peak when the carrier frequency is low (Hu et al., 2017), although this did not allow us to fully rule out adaptation as a key mechanism as the auditory system is tonotopically organised and neuron parameters in low frequency regions may be different to those at higher frequencies. Our model predicts that higher sound levels lead to a preference for earlier onsets (and vice versa), which is consistent with preliminary unpublished observations (Dietz, personal communication).

We then considered mechanisms arising at the level of populations of neurons after binaural integration (for example in the inferior colliculus or auditory cortex). We found that we could again very precisely reproduce the quantitative effect, this time assuming only a local interaction of excitation and inhibition, a lateral inhibition motif that is found in many areas of the brain. This model was

bioRxiv preprint doi: https://doi.org/10.1101/687178; this version posted September 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

also very robust, and was able to reproduce the shift towards the peak at low frequencies based only on the difference in distribution of the interaural phase difference preferences of binaural neurons observed in mammals (McAlpine et al., 2001). This model required that its inputs show a strong synchronisation to the envelope of the stimulus, which can be the result of the onset mechanisms discussed above for example, suggesting that single neuron and population mechanisms may work in concert. The model finally predicts that although lowering the carrier frequency below 500 Hz leads to a later preference (a shift towards the peak), increasing the carrier frequency above 500 Hz should not lead to an earlier preference.

Although some of the mechanisms reviewed in this study have been shown to explain some aspects of the precedence effect (Hartung and Trahiotis, 2001; Braasch and Blauert, 2003; Xia and Shinn-Cunningham, 2011; Xia et al., 2010), none of them were previously tested quantitatively in the context of the experiments by Dietz et al. (2013). Additionally, the use of simple models allows for a comprehensive exploration of some aspects of the precedence effect such as the extraction of IPDs in low-frequency amplitude modulated sounds.

The mechanisms we used in these models are not unique to the auditory system. Our adaptation and onset models are similar or equivalent to membrane potential and synapse dynamics that are found throughout the nervous system (see Methods and Appendix). Adaptation allows neurons to minimize energy costs while encoding persistent stimuli (Jones et al., 2015) and to increase their sensitivity to new stimuli (Fairhall et al., 2001). At the population level, excitatory-inhibitory networks are often found in other sensory modalities, such as in orientation tuning in vision (Ben-Yishai et al., 1995) or in the detection of touch in the field of somatosensation (Mountcastle, 1959). In the case of the auditory system, these circuits have not been positively identified, but would likely be located in the inferior colliculus, where many inhibition circuits exist (Pollak et al., 2011), or higher up in the auditory pathway.

In order to thoroughly investigate the parameter spaces of our models, we abstracted and simplified them as much as possible. This has the great advantage of allowing us to precisely understand the interactions of the different mechanisms, but has the disadvantage of making it difficult to apply the model to other more complex stimuli. Incorporating additional mechanisms such as cochlear and modulation filterbanks would be straightforward but would introduce a large number of additional parameters.

In summary, we found that we could account for the experimental data across a wide range of parameters using multiple single neuron or neural population mechanisms that are not specific to the binaural system, or even to the auditory system. Indeed, it appears that a diverse set of neural mechanisms located in the auditory pathway all emphasize the extraction of the ITD information in the onset of the envelope. While we cannot definitively say which mechanisms are present and contribute to the precedence effect, our results suggest that all the mechanisms we studied have the capacity to do so. This opens the door for two hypotheses that may drive future work.

The first hypothesis is that the preference for early arriving auditory cues (Dietz et al., 2013, 2014; Hu et al., 2017), and the precedence effect more generally, may not be the result of an adaptation specifically for sound localisation in reverberant environments, but may be the result of an evolutionary exaptation or co-opted adaptation. That is, an instance of the brain using already existing dynamics and mechanisms (adaptation, excitation/inhibition, lateral inhibition) to carry out a new function (that may then subsequently have been refined by further evolutionary pressure). This possibility has been suggested more generally under various names, including circuit motifs (Braganza and Beck, 2018), computational primitives (Marcus et al., 2014) and canonical computations (Kouh and Poggio, 2008). In terms of our efforts to try to understand the incredible and general abilities of the brain, this might be an even more exciting possibility than the alternative that there are mechanisms adapted specifically for processing reverberant sounds.

The second hypothesis is that, since the range of parameters equally able to account for the data is so wide, there may be individual differences. These individual differences could come from two sources. It could be that there are a range of neural parameters that all give rise to the same behaviour, as in the case of the pyloric network of the crustacean stomatogastric ganglion (Prinz et al., 2004). Or, it could be that individuals learn different sound localization strategies, making use of the different neural mechanisms that are available to them, in line with the finding that human listeners exhibiting asymmetric hearing can take advantage of multiple adaptive mechanisms to improve their ability to localize sounds (Keating et al., 2016).

bioRxiv preprint doi: https://doi.org/10.1101/687178; this version posted September 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

We therefore strongly encourage researchers to investigate the extent to which new experimental results can be reproduced using simple and general models, and thoroughly explore the parameter space in detail along the lines suggested by O'Leary et al. (2015) and used previously in the auditory system in Goodman et al. (2017).

Methods

Models were implemented in Python using the Brian simulator (Goodman and Brette, 2008, 2013; Stimberg et al., 2019) and the scientific computing packages NumPy and SciPy (Jones et al., 2016). All the code is available online (`https://github.com/neural-reckoning/simple_ambb_modelling`) and can be edited and run directly in the browser via live, interactive notebooks.

Stimulus

The amplitude modulated binaural beats (AMBB) stimuli, designed by Dietz et al. (2013), are amplitude modulated tones whose envelopes follow the equation

$$E(t) = \tfrac{1}{2}(1 - \cos(2\pi f_m t)), \tag{1}$$

where f_m is the modulation frequency of the stimulus. The main characteristic of the AMBB stimulus is its dynamic interaural phase difference (IPD) varying from 0° to 360° in one modulation cycle. This is achieved by introducing an interaural frequency difference between the left and right channels of the sound:

$$S_L(t) = \sin(2\pi(f_c + f_m/2)t)\ E(t) \tag{2}$$
$$S_R(t) = \sin(2\pi(f_c - f_m/2)t)\ E(t). \tag{3}$$

In these equations, f_c refers to the carrier frequency while S_L and S_R represent respectively the left and right channels. In the first part of section Results (up until section Homogeneous neurons without onset mechanisms do not exhibit an early preference) we study peripheral adaptation mechanisms which are known to be monaural, and therefore only use the envelope of the stimulus as input. Later, while investigating the influence of the carrier frequency (section Homogeneous neurons without onset mechanisms do not exhibit an early preference) we add a carrier channel at 500 Hz.

$$S(t) = \cos(2\pi f_c t)\ E(t). \tag{4}$$

Single neuron model

We start off by describing the modelling of single cells in the auditory periphery. In this study, cochlear filtering was not necessary as all signals were narrowband. We model amplification, compression and inner hair cell dynamics in a fairly standard way as follows. First we half-wave rectify and compress the signal as

$$A_{\text{pre}}(t) = ([10^{L/20} S(t)]^+)^\gamma, \tag{5}$$

where the operator $[]^+$ represents half-wave rectification ($[x]^+ = x$ when $x > 0$ otherwise $[x]^+ = 0$), L is the gain (in dB) and γ is the compression ($\gamma = 1$ is no compression, and $\gamma = 1/3$ is a commonly used value). The signal is then passed through a first order low pass filter with time constant τ_{ihc}:

$$\tau_{\text{ihc}} \frac{\mathrm{d}A}{\mathrm{d}t} = A_{\text{pre}} - A. \tag{6}$$

The resulting signal was then passed to an adaptation stage inspired by the functioning of the synapse between the inner hair cells and the auditory nerve fibers. This model is a simplification of published adaptation models (Meddis, 1986, 1988; Westerman and Smith, 1988; Zhang and Carney, 2005) to a single reservoir and consequently a single time-constant. The quantity of available neurotransmitter Q (between 0 and 1) is depleted at a rate proportional to the product of the incoming signal and the remaining amount of neurotransmitter (κQA), and replenished at a rate

bioRxiv preprint doi: https://doi.org/10.1101/687178; this version posted September 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

proportional to the deficit ($\rho(1-Q)$). The output of the model R_a is the product of the input signal A with the instantaneous quantity of available neurotransmitters Q and models the resulting firing rate of the nerve fibers:

$$\frac{\mathrm{d}Q}{\mathrm{d}t} = -\kappa QA + \rho(1-Q) \tag{7}$$

$$R_a = AQ. \tag{8}$$

To give an idea of the behaviour of this model, with a fixed input $A = 1$ it decays exponentially to a value $1-\alpha = \rho/(\kappa+\rho)$ with time constant $\tau_a = 1/(\kappa+\rho)$. We can invert this ($\kappa = \alpha/\tau_a$, $\rho = (1-\alpha)/\tau_a$) and use α and τ_a as parameters, with α representing the strength of adaptation ($\alpha = 0$ no adaptation, α close to 1 for strong adaptation) and τ_a the time constant of adaptation (for a fixed reference signal).

Finally, we use the following equations to model either onset cells or excitatory and inhibitory dynamics:

$$\tau_e \frac{\mathrm{d}R_e}{\mathrm{d}t} = R_a - R_e \tag{9}$$

$$\tau_i \frac{\mathrm{d}R_i}{\mathrm{d}t} = R_a - R_i \tag{10}$$

$$R = [R_e - \beta R_i]^+. \tag{11}$$

Each current was produced by low-pass filtering the signal with different time-constants τ_e and τ_i. If $\tau_e < \tau_i$ are close this acts as a differentiator.

The model is run for a number of cycles (at least 1) until it settles into a periodic state, and then the IPD returned by the model is the phase corresponding to the time when the output signal $R(t)$ is at a maximum on the next cycle:

$$\text{extracted phase} = 2\pi f_m \cdot \arg\max_t R(t)\ . \tag{12}$$

Without any additional mechanisms (adaptation or onset, i.e. $\alpha = \beta = 0$), the IPD returned by the model is consistently equal to 180°.

Error measures

To evaluate the fit of the model to the data, we used the maximum absolute error in the IPD across modulation frequencies:

$$\varepsilon = \max_{f_m} |IPD_{\text{data}}(f_m) - IPD_{\text{model}}(f_m)|. \tag{13}$$

The same error measure was also used in the population model. In cases where we compare errors at $f_c = 200$ Hz and $f_c = 500$ Hz we compute ε_{200} and ε_{500} using the equation above, and compute a total error based on a weighted maximum of these two:

$$\varepsilon = \max\{\ \varepsilon_{500},\ \tfrac{1}{3}\varepsilon_{200}\ \}. \tag{14}$$

We weight the error at $f_c = 200$ Hz lower as we do not have direct and equivalent experimental data at this carrier frequency (Hu et al., 2017).

Modulation measures

We computed several measures of the variation of phase-locking and firing rates with modulation frequency.

The mean *temporal modulation transfer function (tMTF)* measures the degree to which the model

bioRxiv preprint doi: https://doi.org/10.1101/687178; this version posted September 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

phase-locks to the envelope across modulation frequencies, and is computed as:

$$\phi(t) = 2\pi f_m t \tag{15}$$

$$\overline{R(f_m)} = \frac{1}{2\pi}\int_0^{2\pi} R(f_m, \phi(t))\, dt \tag{16}$$

$$v(f_m) = \frac{2\pi}{\overline{R(f_m)}}\sqrt{\left(\int_0^{2\pi} R(f_m, \phi(t))\sin(\phi(t))\, dt\right)^2 + \left(\int_0^{2\pi} R(f_m, \phi(t))\cos(\phi(t))\, dt\right)^2} \tag{17}$$

$$tMTF = \frac{1}{N_{f_m}}\sum_{f_m} v(f_m) \tag{18}$$

Here $\phi(t)$ is the modulation phase at time t, $\overline{R(f_m)}$ is the mean firing rate across one modulation cycle, and $v(f_m)$ is the vector strength.

The *rate modulation transfer function*, $rMTF(f_m)$, is the ratio of the mean firing rate at frequency f_m to the maximum across all f_m, and the mean rMTF is the mean of this over all f_m, so that a low value indicates a variable rMTF and a value near 1 indicates a fixed rMTF:

$$rMTF(f_m) = \frac{\overline{R(f_m)}}{\max_{f_m} \overline{R(f_m)}} \tag{19}$$

$$rMTF = \frac{1}{N_{f_m}}\sum_{f_m} rMTF(f_m) \tag{20}$$

In the set of equations above and in the rest of this method section, a line over a symbol represents the operation of taking a mean over time t.

The *temporal modulation depth (tMD)* is computed as the normalised difference between the minimum and maximum vector strength across modulation frequencies:

$$tMD = \max_{f_m} v(f_m) - \min_{f_m} v(f_m) \tag{21}$$

The *rate modulation depth (RMD)* is similarly computed as the normalised difference between the minimum and maximum firing rates across modulation frequencies:

$$rMD = 1 - \frac{\min_{f_m} \overline{R(f_m)}}{\max_{f_m} \overline{R_(f_m)}} \tag{22}$$

The *temporal best modulation frequency (tBMF)* is the modulation frequency maximizing the vector strength:

$$tBMF = \arg\max_{f_m} v(f_m) \tag{23}$$

The *rate best modulation frequency (rBMF)* is the modulation frequency maximizing the mean firing rate:

$$rBMF = \arg\max_{f_m} \overline{R(f_m)} \tag{24}$$

Population model

We model the experiment of Dietz et al. (2013) by comparing the similarity of the neural population response to the AMBB stimulus to the response to an amplitude modulated tone with a static IPD. The neurons in these populations are binaural and IPD-tuned, and interact with each other by lateral inhibition. We explain each of these elements below.

Each of the binaural neurons in the first layer is sensitive to IPD (ϕ), with a tuning curve centered around its best IPD (BIPD $\tilde{\phi}$) given by

$$T(\tilde{\phi}, \phi) = \left(\cos\frac{\tilde{\phi} - \phi}{2}\right)^k. \tag{25}$$

bioRxiv preprint doi: https://doi.org/10.1101/687178; this version posted September 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

This tuning curve has previously been shown to fit various animal models for different values of k (Harper and McAlpine, 2004; Joris et al., 2006; Fischer et al., 2008; Goodman et al., 2013) and is shown in figure 10C. We next assume that the response is proportional to the stimulus amplitude. For the AMBB stimulus, the amplitude at time t is given by $E(t)$ as above, and the IPD $\phi(t) = 2\pi f_m t$, so the response of the neuron with BIPD $\tilde{\phi}$ at time t is

$$F(\tilde{\phi}, t) = T(\tilde{\phi}, \phi(t)) \cdot E(t). \tag{26}$$

The original amplitude modulated signal has modulation depth 1 (the envelope goes from 0 to 1), however the modulation depth of the neural response to this stimulus may differ, and may have a higher or lower synchronisation or vector strength with respect to the modulation frequency. We model this by replacing the envelope $E(t)$ above with

$$E(t) = \tfrac{1}{2}[1 - m\cos\phi(t)]^+. \tag{27}$$

The variable m represents the synchronization of the neural response to the envelope of the stimulus. When $m = 0$ there is no synchronization of the neural response to the stimulus. When $m = 1$ the synchronization is the same as the modulation depth of the stimulus. Finally, when $m > 1$, the neuron exhibits enhanced synchronization (as seen for example in onset cells).

For computational efficiency, we discretize the model by choosing $N_{\text{BIPD}} = 100$ BIPDs uniformly between 0° and 360°, and (for each modulation frequency) $N_t = 250$ time steps per modulation cycle (so the sample width is $1/N_t f_m$ at modulation frequency f_m). We write the discrete BIPDs and times as $\tilde{\phi}_i = 2\pi i/N_{\text{BIPD}}$ and $t_j = j/N_t f_m$. Note that this does not mean we consider BIPDs to be necessarily uniformly distributed: below we will weight the responses according to the density distribution of BIPDs. With these discretizations, we can now write the model in matrix form as follows:

$$T_{ij} = T(\tilde{\phi}_i, \phi(t_j)) \tag{28}$$

$$E_{ij} = \begin{cases} E(t_i) & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} \tag{29}$$

$$F = TE. \tag{30}$$

The pattern of activity generated by equation (30) can be seen in Figure 10A.

BIPD distribution

We used an experimentally measured BIPD distribution, with the mean μ and the standard deviation σ of the BIPDs measured from guinea pigs in McAlpine et al. (2001). The density of BIPDs in our model then followed a bimodal normal distribution based on these statistics. The density function of this distribution is proportional to:

$$d(\tilde{\phi}) = e^{-(\mu-\tilde{\phi})^2/2\sigma^2} + e^{-(-\mu-\tilde{\phi})^2/2\sigma^2}. \tag{31}$$

At $f_c = 500$ Hz, this distribution has values close to 0 at BIPDs around 180° and exhibits two peaks respectively around IPDs equal to 45° and 315°. BIPD distributions at different frequencies can be seen in figure 10B.

To incorporate the non-uniform BIPD distribution into the uniform matrix formulation above, we multiply the response F_{ij} with the density $d(\tilde{\phi}_i)$ (figure 10A). The model here is continuous valued, but can be considered to represent mean firing rates or spike counts. Multiplying by the density therefore makes the response proportionate to the mean number of spikes from neurons tuned to a given BIPD. As a consequence, some of the weighted responses with $\tilde{\phi}$ close to 180° will be close to 0 because there are next to no neurons with BIPD close to 180°. We write the response of the weighted network as F_W, in matrix form:

$$D_{ij} = \begin{cases} d(\tilde{\phi}_i) & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} \tag{32}$$

$$F_W = DF = DTE. \tag{33}$$

bioRxiv preprint doi: https://doi.org/10.1101/687178; this version posted September 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

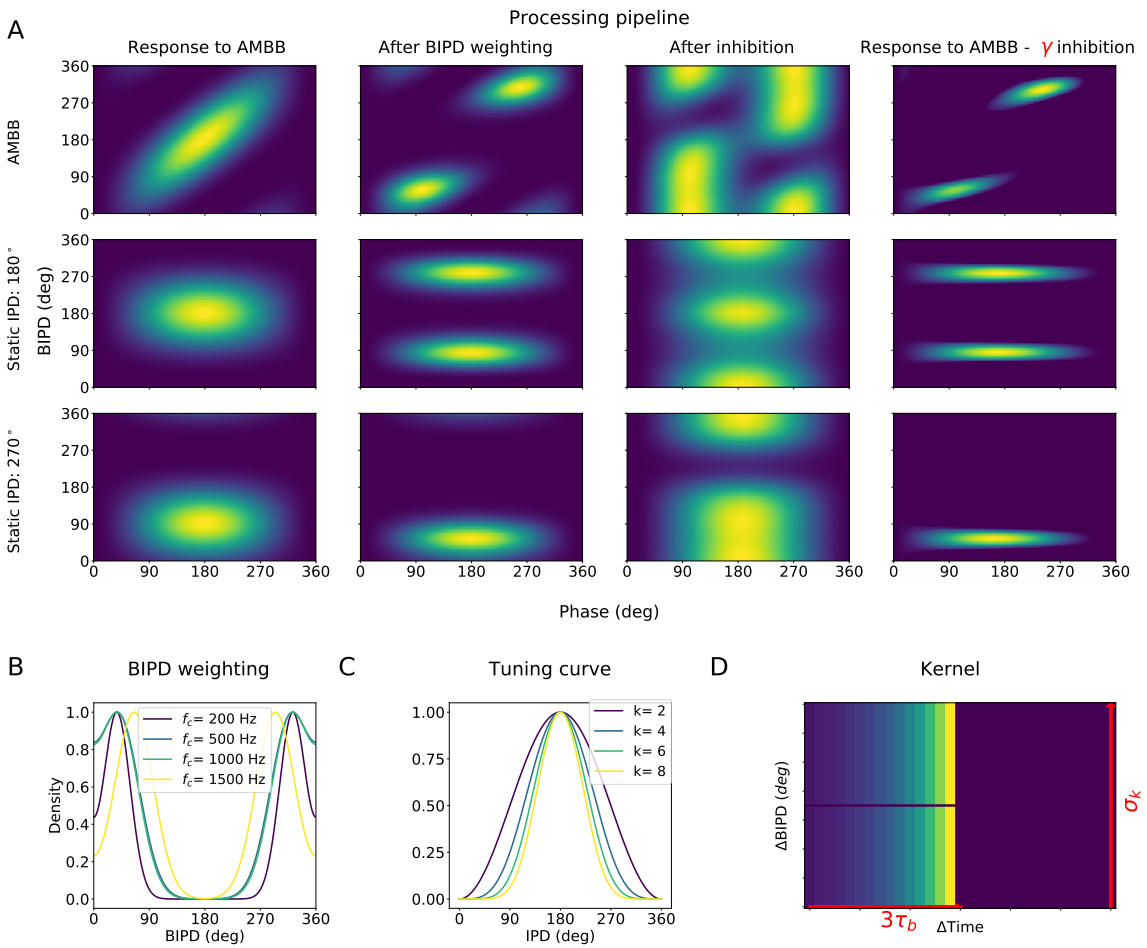

Figure 10: (A) Top row: the first column shows the response of the neurons to the stimulus. The response of the network after BIPD weighting is shown in the second column. The third column shows the inhibition layer, i.e. the network after inhibition. The last column shows the response of the network after subtracting the γ weighted inhibition layer from the BIPD weighted neural activity. Second row: Same as first row but when the stimulus is an AM tone with a static IPD=180°. Third row: same but with IPD=90°. (B) Shape of the BIPD distribution at different carrier frequencies. (C) Shape of the tuning curve centered around $BIPD = 180°$ with different values of the exponent k. (D) Shape of the inhibition kernel. The weight of inhibition decreases exponentially with time following the time constant τ_b. The fraction of neurons involved in the inhibition process is set by the variable σ_k.

Lateral inhibition

The next layer of neurons in the network have a lateral inhibition mechanism. This is modelled with a synaptic connectivity assuming a narrowly tuned excitation and broad lateral inhibition (figure 11). We assume that excitatory neurons are fast compared to inhibitory neurons, which we model by applying a low-pass filter to the inhibitory neurons (approximating a number of neural processes such as membrane potential or synapse dynamics).

The individual inhibitory currents $i_{\text{inh}}(\tilde{\phi}, t)$ are found by low-pass filtering the response. This can be represented as a convolution of the response with an exponentially decaying kernel:

$$i_{\text{inh}}(\tilde{\phi}, t) = \int_{-\infty}^{0} e^{\tau/\tau_b} F_W(\tilde{\phi}, \tau - t)\, \mathrm{d}\tau. \tag{34}$$

The second layer of neural responses is then formed by taking the first layer as the narrowly tuned excitatory response, and subtracting the inhibitory currents (weighted by a factor γ) for all BIPDs within a window of size σ_k. The variable σ_k is the fraction of the set of BIPDs involved in the inhibition, so $\sigma_k = 0$ represents no inhibition and $\sigma_k = 1$ represents all neurons receiving the same inhibition proportional to the sum of the population activity. The summed inhibitory current is

bioRxiv preprint doi: https://doi.org/10.1101/687178; this version posted September 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

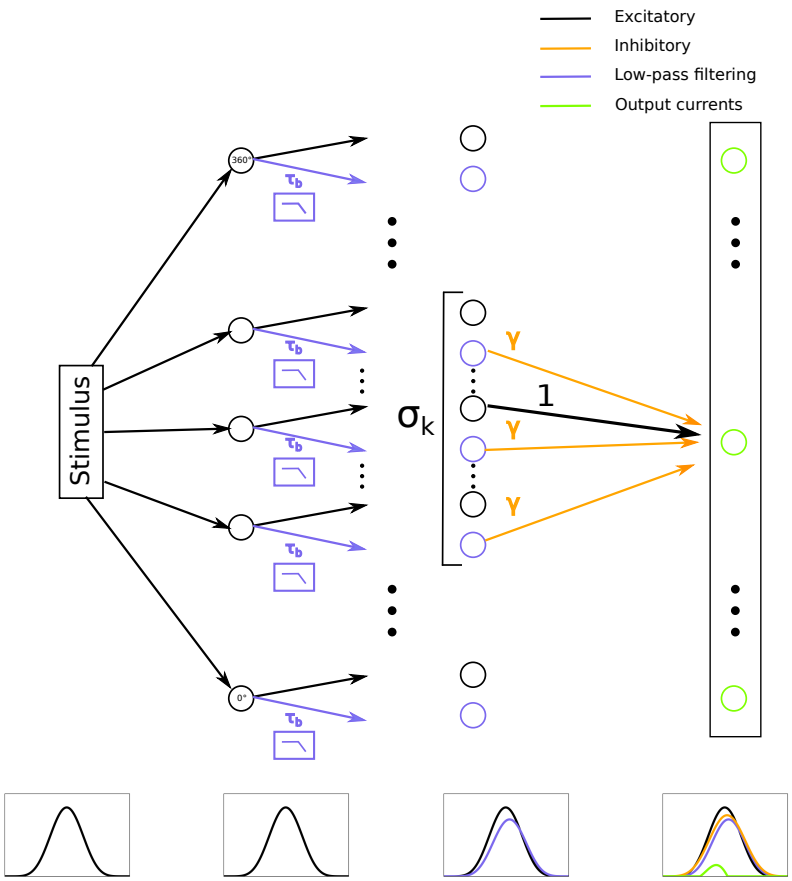

Figure 11: Structure of the network. Black circles represent the initial response of the neurons in the network to the AMBB stimulus. Each neuron is tuned to different BIPDs, from 0° to 360°. Blue circles represent the local inhibitory currents, and are low-pass filtered version of the black circles. The resulting inhibition currents, after low-pass filtering, are weighted by a factor γ and summed within the BIPD window of width σ_k. Green circles represent the output neural activity for each BIPD. Each green circle is the response of the neuron to the stimulus (the black circles) minus the weighted sum of the inhibitory currents (orange arrows). Pattern-match decoding is performed on the pattern of currents represented by the green circles. Curves below the schematic represent the exaggerated output of the network at each stage of processing. Line colours correspond to the network elements above.

given by

$$I_{\text{inh}}(\tilde{\phi}, t) = \sum_{\psi \in S_k(\tilde{\phi})} i_{inh}(\psi, t) = \sum_{\psi \in S_k(\tilde{\phi})} \int_{-\infty}^{0} e^{s/\tau_b} F_W(\psi, t-s) \mathrm{d}s, \tag{35}$$

where $S_k(\tilde{\phi})$ is the set of BIPDs within a distance σ_k of $\tilde{\phi}$. The net response of the second layer is therefore the excitatory response minus γ times the summed inhibitory current:

$$F_{WI}(\tilde{\phi}, t) = [F_W(\tilde{\phi}, t) - \gamma I_{\text{inh}}(\tilde{\phi}, t)]^+. \tag{36}$$

From equations 34 and 35, we can see that the summed inhibitory current I_{inh} can be written as the 2D convolution of the response of the network with a kernel K. In other words, equation (36) can be rewritten as

$$F_{WI} = [F_W - \gamma F_W \circledast K]^+, \tag{37}$$

bioRxiv preprint doi: https://doi.org/10.1101/687178; this version posted September 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

where $\circledast$ represents 2D convolution. The kernel K is the outer product of K_t and K_ϕ, defined as

$$K_\phi(\psi) = \begin{cases} 1 & \text{if } |\psi| < 2\pi\sigma_k \\ 0 & \text{otherwise} \end{cases} \tag{38}$$

$$K_t(s) = \begin{cases} e^{s/\tau_b} & \text{if } s < 0 \\ 0 & \text{otherwise} \end{cases} \tag{39}$$

$$K(\psi, s) = K_t(s) \cdot K_\phi(\psi). \tag{40}$$

For the discretized formulation, we approximate K by cutting off $s < -3\tau_b$. With this approximation made, equation 37 becomes a matrix equation (although note that some care is needed to handle the circular boundaries in the convolution). This matrix formulation was essential in optimising the model to run in a reasonable time and allowing us to evaluate such a large number of parameter sets.

The result of the convolution process is shown in the fourth column of figure 10A. An important characteristic of the network's response after lateral inhibition is the introduction of phase-shifts in the neuronal response. The distribution of these phase-shifts mostly depend on the value of τ_b and on the modulation frequency (figure 12A and 12B) and becomes critical during the similarity scoring process.

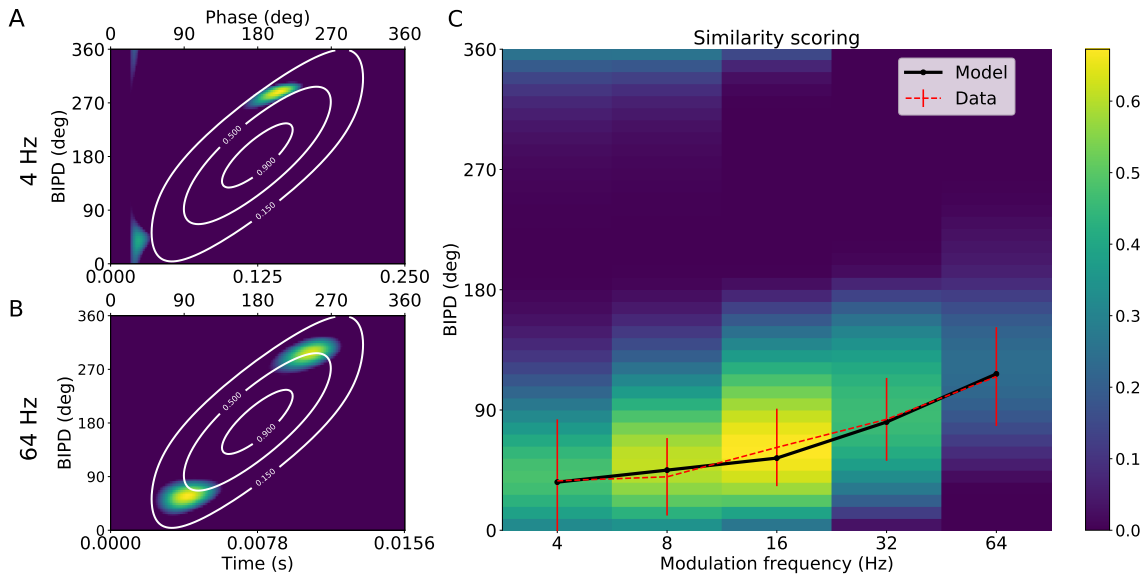

Figure 12: (A) Response of the network to the modelled amplitude modulated binaural beat (AMBB) stimulus with modulation frequency $f_m = 4$ Hz. Response of the network without inhibition is shown in white contours. (B) Same as A but with $f_m = 64$ Hz. (C) Similarity scoring. Similarity map showing similarity values for each modulation frequency, between the response of the full model to AMBB and the response to the amplitude modulated tones with static interaural phase differences.

Similarity scoring

The experiment described in Dietz et al. (2013) relies on the subjects matching static IPD pointers to the location they perceived while listening to the AMBB stimulus. To model this process, we used a similarity scoring method between the response produced by the AMBB stimulus and the response produced by different static IPD pointers. We initially considered the hemispheric decoder (Stecker et al., 2005; Grothe et al., 2010) but we were not able to use it in the context of this experiment because of its mono-dimensionality. Indeed, if the hemispheric ratio in response to the AMBB stimulus as a function of phase difference is $R(\phi)$ then R must be 2π-periodic by construction of the experiment. This means that all but at most two values of R must have multiple inverses in $[0, 2\pi)$ and therefore it is impossible to use the hemispheric difference to match the static IPD successfully. Instead, we used a decoder inspired by the pattern match decoder described in Goodman et al. (2013). It works by computing a similarity measure between the response of the network F_{WI} to

bioRxiv preprint doi: https://doi.org/10.1101/687178; this version posted September 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

the AMBB stimulus and the response of the network to an amplitude modulated tone with a static IPD $S_{WI}(\tilde{\phi})$. The similarity was scored over a full stimulus cycle. 40 static IPDs were selected, linearly distributed between 0° and 360°. Examples of such patterns can be found in figure 10A second and third row. We score the similarity between the response to the AMBB stimulus and the static IPD stimuli using the dot product between the flattened arrays F_{WI} and $S_{WI}(\tilde{\phi})$. The IPD maximizing the similarity scoring was returned by the model (figure 12C). This is equivalent to the cosine similarity metric as $|S_{WI}(\tilde{\phi})|$ is the same for all $\tilde{\phi}$.

$$\phi_{\text{out}} = \arg\max_{\tilde{\phi}} < S_{WI}(\tilde{\phi}), F_{WI} > \tag{41}$$

Starting phase

Simulations were run with 8 different starting IPDs (ϕ_{start}) between 0° and 360° as in Dietz et al. (2013). ϕ_{start} corresponds to the value of the instantaneous IPD at the onset of the stimulus (when $t = 0$). In order to compare to the curves in Dietz et al. (2013), model curves were plotted using the circular mean of the extracted IPDs over the 8 different ϕ_{start}.

Acknowledgements

We would like to thank Mathias Dietz, David McAlpine and Andrew Brughera for fascinating discussions of this work and comments on the paper.

References

Ben-Yishai, R., R. L. Bar-Or, and H. Sompolinsky (1995). Theory of orientation tuning in visual cortex. *Proceedings of the National Academy of Sciences 92*(9), 3844–3848.

Blauert, J. (1997). *Spatial hearing: the psychophysics of human sound localization*. MIT press.

Braasch, J. and J. Blauert (2003). The precedence effect for noise bursts of different bandwidths. ii. comparison of model algorithms. *Acoustical Science and Technology 24*(5), 293–303.

Braganza, O. and H. Beck (2018, mar). The Circuit Motif as a Conceptual Tool for Multilevel Neuroscience. *Trends in neurosciences 41*(3), 128–136.

Brown, A. D., G. C. Stecker, and D. J. Tollin (2015). The precedence effect in sound localization. *Journal of the Association for Research in Otolaryngology 16*(1), 1–28.

Cherry, C. E. (1953). Some experiments on the recognition of speech, with one and with two ears. *J Acoust Soc Am 25*(5), 975–979.

Dietz, M., T. Marquardt, N. H. Salminen, and D. McAlpine (2013). Emphasis of spatial cues in the temporal fine structure during the rising segments of amplitude-modulated sounds. *Proceedings of the National Academy of Sciences of the United States of America 110*(37), 15151–6.

Dietz, M., T. Marquardt, A. Stange, M. Pecka, B. Grothe, and D. McAlpine (2014). Emphasis of spatial cues in the temporal fine structure during the rising segments of amplitude-modulated sounds II: single-neuron recordings. *Journal of neurophysiology 111*(10), 1973–85.

Fairhall, A. L., G. D. Lewen, W. Bialek, and R. R. d. R. van Steveninck (2001). Efficiency and ambiguity in an adaptive neural code. *Nature 412*(6849), 787.

Ferragamo, M. J. and D. Oertel (2002). Octopus cells of the mammalian ventral cochlear nucleus sense the rate of depolarization. *J Neurophysiol 87*(5), 2262–2270.

Fischer, B. J., G. B. Christianson, and J. L. Pena (2008). Cross-Correlation in the Auditory Coincidence Detectors of Owls. *Journal of Neuroscience 28*(32), 8107–8115.

bioRxiv preprint doi: https://doi.org/10.1101/687178; this version posted September 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Freyman, R. L., K. S. Helfer, D. D. McCall, and R. K. Clifton (1999). The role of perceived spatial separation in the unmasking of speech. *The Journal of the Acoustical Society of America 106*(6), 3578–3588.

Goodman, D. and R. Brette (2008). Brian: a simulator for spiking neural networks in python. *Frontiers in neuroinformatics 2*, 5.

Goodman, D. and R. Brette (2013). Brian simulator. *Scholarpedia 8*(1), 10883.

Goodman, D., I. Winter, A. Léger, A. de Cheveigné, and C. Lorenzi (2017). Modelling firing regularity in the ventral cochlear nucleus: Mechanisms, and effects of stimulus level and synaptopathy. *Hearing Research*.

Goodman, D. F., V. Benichoux, and R. Brette (2013). Decoding neural responses to temporal cues for sound localization. *Elife 2*, e01312.

Grothe, B., M. Pecka, and D. McAlpine (2010). Mechanisms of sound localization in mammals. *Physiological reviews 90*(3), 983–1012.

Harper, N. S. and D. McAlpine (2004). Optimal neural population coding of an auditory spatial cue. *Nature 430*(7000), 682.

Hartung, K. and C. Trahiotis (2001). Peripheral auditory processing and investigations of the "precedence effect" which utilize successive transient stimuli. *The Journal of the Acoustical Society of America 110*(3), 1505–1513.

Hu, H., S. D. Ewert, D. McAlpine, and M. Dietz (2017, mar). Differences in the temporal course of interaural time difference sensitivity between acoustic and electric hearing in amplitude modulated stimuli. *The Journal of the Acoustical Society of America 141*(3), 1862–1873.

Humphries, C., E. Liebenthal, and J. R. Binder (2010). Tonotopic organization of human auditory cortex. *Neuroimage 50*(3), 1202–1211.

Jones, D. L., E. C. Johnson, and R. Ratnam (2015). A stimulus-dependent spike threshold is an optimal neural coder. *Frontiers in computational neuroscience 9*, 61.

Jones, E., T. Oliphant, P. Peterson, et al. (2016). Scipy: Open source scientific tools for python, 2001.

Joris, P. X., B. Van de Sande, D. H. Louage, and M. van der Heijden (2006). Binaural and cochlear disparities. *Proceedings of the National Academy of Sciences*.

Keating, P., O. Rosenior-Patten, J. C. Dahmen, O. Bell, and A. J. King (2016). Behavioral training promotes multiple adaptive processes following acute hearing loss. *Elife 5*, e12264.

Kouh, M. and T. Poggio (2008, jun). A Canonical Neural Circuit for Cortical Nonlinear Operations. *Neural Computation 20*(6), 1427–1451.

Lindemann, W. (1986a). Extension of a binaural cross-correlation model by contralateral inhibition. i. simulation of lateralization for stationary signals. *The Journal of the Acoustical Society of America 80*(6), 1608–1622.

Lindemann, W. (1986b). Extension of a binaural cross-correlation model by contralateral inhibition. ii. the law of the first wave front. *The Journal of the Acoustical Society of America 80*(6), 1623–1630.

Litovsky, R. Y., H. S. Colburn, W. A. Yost, and S. J. Guzman (1999). The precedence effect. *The Journal of the Acoustical Society of America 106*(4), 1633–1654.

Marcus, G., A. Marblestone, and T. Dean (2014). The atoms of neural computation.

McAlpine, D., D. Jiang, and A. R. Palmer (2001). A neural code for low-frequency sound localization in mammals. *Nature neuroscience 4*(4), 396.

Meddis, R. (1986). Simulation of mechanical to neural transduction in the auditory receptor. *The Journal of the Acoustical Society of America*.

bioRxiv preprint doi: https://doi.org/10.1101/687178; this version posted September 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Meddis, R. (1988). Simulation of auditory–neural transduction: Further studies. *The Journal of the Acoustical Society of America 83*(3), 1056–1063.

Mountcastle, V. (1959). Tps: Neural mechanisms subserving cutaneus sensitivity, with special reference to the role of afferent inhibition in sensory perception and discrimination. *Johns Hopkins Med. J. 105*, 201–232.

O'Leary, T., A. C. Sutton, and E. Marder (2015). Computational models in the age of large datasets. *Current Opinion in Neurobiology 32*, 87–94.

Platkiewicz, J. and R. Brette (2010, jul). A threshold equation for action potential initiation. *PLoS Computational Biology 6*(7), e1000850.

Platkiewicz, J. and R. Brette (2011, may). Impact of Fast Sodium Channel Inactivation on Spike Threshold Dynamics and Synaptic Integration. *PLoS Comput Biol 7*(5), e1001129.

Pollak, G. D., R. Xie, J. X. Gittelman, S. Andoni, and N. Li (2011). The dominance of inhibition in the inferior colliculus. *Hearing research 274*(1-2), 27–39.

Prinz, A. A., D. Bucher, and E. Marder (2004). Similar network activity from disparate circuit parameters. *Nature neuroscience 7*(12), 1345–1352.

Ress, D. and B. Chandrasekaran (2013). Tonotopic organization in the depth of human inferior colliculus. *Frontiers in human neuroscience 7*, 586.

Rothman, J. S. and P. B. Manis (2003). The roles potassium currents play in regulating the electrical activity of ventral cochlear nucleus neurons. *Journal of neurophysiology 89*(6), 3097–3113.

Russell, I. J. and P. M. Sellick (1983, may). Low-frequency characteristics of intracellularly recorded receptor potentials in guinea-pig cochlear hair cells. *The Journal of Physiology 338*(1), 179–206.

Spencer, M. J., D. B. Grayden, I. C. Bruce, H. Meffin, and A. N. Burkitt (2012). An investigation of dendritic delay in octopus cells of the mammalian cochlear nucleus. *Frontiers in Computational Neuroscience 6*(October), 1–19.

Spencer, M. J., H. Meffin, A. N. Burkitt, and D. B. Grayden (2018, jun). Compensation for Traveling Wave Delay Through Selection of Dendritic Delays Using Spike-Timing-Dependent Plasticity in a Model of the Auditory Brainstem. *Frontiers in Computational Neuroscience 12*, 36.

Stecker, G. C., I. A. Harrington, and J. C. Middlebrooks (2005). Location coding by opponent neural populations in the auditory cortex. *PLoS biology 3*(3), e78.

Stimberg, M., R. Brette, and D. F. M. Goodman (2019, August). Brian 2, an intuitive and efficient neural simulator. *eLife 8*, e47314.

Tollin, D. (1998). Conputational model of the lateralisation of clicks and their echoes. *Proc. the NATO Advanced Study Institute on Computational Hearing*, 77–82.

Tsodyks, M., K. Pawelzik, and H. Markram (1998). Neural Networks with Dynamic Synapses. *Neural Computation*.

Tsodyks, M. and S. Wu (2013, oct). Short-term synaptic plasticity. *Scholarpedia 8*(10), 3153.

Wallach, H., E. B. Newman, and M. R. Rosenzweig (1949). A precedence effect in sound localization. *The Journal of the Acoustical Society of America 21*(4), 468–468.

Westerman, L. A. and R. L. Smith (1988). A diffusion model of the transient response of the cochlear inner hair cell synapse. *J. Acoust. Soc. Am. 83*(June), 2266–2276.

Xia, J., A. Brughera, H. S. Colburn, and B. Shinn-Cunningham (2010). Physiological and psychophysical modeling of the precedence effect. *Journal of the Association for Research in Otolaryngology 11*(3), 495–513.

Xia, J. and B. Shinn-Cunningham (2011). Isolating mechanisms that influence measures of the precedence effect: Theoretical predictions and behavioral tests. *The Journal of the Acoustical Society of America 130*(2), 866–882.

bioRxiv preprint doi: https://doi.org/10.1101/687178; this version posted September 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Zhang, X. and L. H. Carney (2005). Analysis of models for the synapse between the inner hair cell and the auditory nerve. *The Journal of the Acoustical Society of America 118*(3), 1540.

Zurek, P. M. (1987). The precedence effect. In *Directional hearing*, pp. 85–105. Springer.

Appendix

Onset cell is a special case of excitation/inhibition model

In the single neuron onset model (equations 9-11), R_e could represent an excitatory pathway, R_i an inhibitory pathway, and $R = [R_e - \beta R_i]^+$ the output of excitation minus inhibition with a relative inhibitory strength β. However, if $\beta = 1$ and $\tau_e \ll \tau_i$ the same equations can also be interpreted as a single onset neuron. The equations are approximately equal to a rate-based version of the octopus cell model of Spencer et al. (2018), which is itself a simplification of the models of Spencer et al. (2012) and Ferragamo and Oertel (2002). In their spiking model,

$$\begin{aligned} \tau_v \frac{\mathrm{d}v}{\mathrm{d}t} &= x - v \\ \tau_x \frac{\mathrm{d}x}{\mathrm{d}t} &= -x, \end{aligned}$$

where v is the membrane potential, x is a synaptic conductance that is increased by a fixed amount for each incoming spike, and the model fires a spike when $\mathrm{d}v/\mathrm{d}t$ crosses some threshold. In the limit of a large number of incoming spikes with small synaptic weights we can replace the discrete changes in x by a continuous current $I(t)$ to give

$$\tau_x \frac{\mathrm{d}x}{\mathrm{d}t} = I(t) - x.$$

Setting $I(t) = R_a(t)$ and $\tau_x = \tau_e$ we can immediately see that $R_e = x$. Now v evolves with time constant τ_v which is much larger than τ_x so on the time scale of the evolution of v we will have that $x \approx I(t)$ and so

$$\tau_v \frac{\mathrm{d}v}{\mathrm{d}t} \approx I(t) - v.$$

Setting $\tau_v = \tau_e$ this gives us $v \approx R_i$. In the limit of a large number of output spikes, the firing rate of the onset cell will be approximately proportional to

$$\frac{\mathrm{d}v}{\mathrm{d}t} = \frac{x - v}{\tau_v} \approx \frac{R_e - R_i}{\tau_v} = \frac{R}{\tau_v}.$$

In other words, the firing rate is proportional to R.

Analytic solution for $\alpha = 0$

We set up the following set of differential equations:

$$\tau_e \frac{\mathrm{d}R_e}{\mathrm{d}t} = E(t) - R_e \tag{42}$$

$$\tau_i \frac{\mathrm{d}R_i}{\mathrm{d}t} = E(t) - R_i \tag{43}$$

$$R = R_e - \beta R_i, \tag{44}$$

where $E(t)$ is the envelope as above. We solve the differential equations with the following periodic boundary conditions (corresponding to the settled solution after the initial few cycles):

$$R_e(0) = R_e(1/f_m) \tag{45}$$

$$R_i(0) = R_i(1/f_m). \tag{46}$$

$$\tag{47}$$

bioRxiv preprint doi: https://doi.org/10.1101/687178; this version posted September 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Solving these equations gives

$$R_e(t) = \frac{1+\sigma_e^2 - \cos\phi - \sigma_e \sin\phi}{2(1+\sigma_e^2)} \tag{48}$$

$$R_i(t) = \frac{1+\sigma_i^2 - \cos\phi - \sigma_i \sin\phi}{2(1+\sigma_i^2)}, \tag{49}$$

where

$$\phi = 2\pi f_m t \tag{50}$$

$$\sigma_e = 2\pi f_m \tau_e \tag{51}$$

$$\sigma_i = 2\pi f_m \tau_i. \tag{52}$$

We want to find the maximum value of t where $\mathrm{d}R/\mathrm{d}t = 0$ and $\mathrm{d}^2R/\mathrm{d}t^2 < 0$, which we find to be the value of θ so that

$$\cos\theta = \beta(1+\sigma_e^2) - (1+\sigma_i^2) \tag{53}$$

$$\sin\theta = \beta\sigma_i(1+\sigma_e^2) - \sigma_e(1+\sigma_i^2). \tag{54}$$

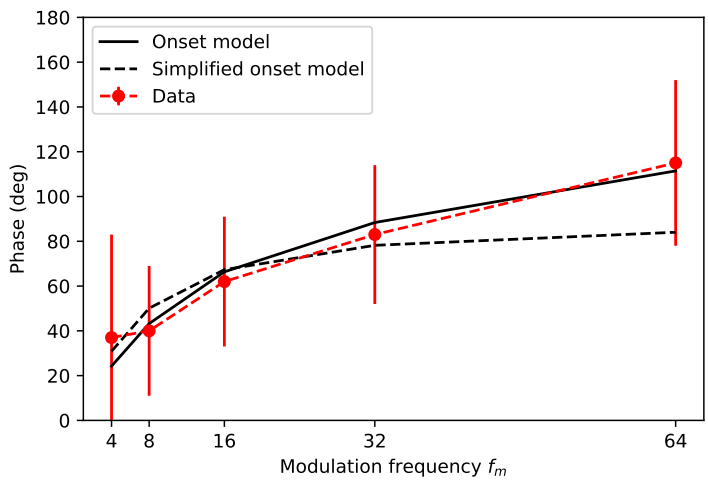

Figure 13: Best fit to data for analytically solved single neuron onset model.

The parameter θ can be extracted directly as $\theta = \tan^{-1}(\sin\theta/\cos\theta)$ but some care has to be taken to get the correct value of $\theta \in [0, 2\pi)$. This gives us a three parameter equation that can fit the data extremely well (figure 13). If we measure f_m in kHz then the following is a close fit to the data:

$$\theta(f_m) = \tan^{-1}\left(\frac{0.001 - 0.79 f_m^2}{0.1 f_m - 1.63 f_m^3}\right). \tag{55}$$

However, we can simplify further if we assume that $\sigma_e\sigma_i = \sigma_e^2 = \sigma_i^2 = 0$ to get

$$\theta = \tan^{-1}\left(\frac{2\pi(\beta\tau_i - \tau_e)}{\beta - 1} \cdot f_m\right) = \tan^{-1}(A f_m), \tag{56}$$

for a single constant A. Figure 13 shows that this approximation gives a reasonable fit with $A = 0.15$ when f_m is small but breaks down for $f_m = 64$ Hz.