Imperial College London,
Department of Electrical and Electronic Engineering

The role of canonical neural computations in sound localization

Jean-Hugues Lestang

Supervisor: Dr. Dan F.M. Goodman

Submitted in fulfilment of the requirements for the degree of Doctor of Philosophy in Electrical and Electronic Engineering and the Diploma of Imperial College London, October 2019

Abstract

Localizing sounds is an important ability for many species. However, reverberative sounds present a significant challenge to the auditory system as later arriving reverberations may carry confounding localization cues. The 'precedence effect' refers to a set of perceptual behaviours related to this situation. Studies investigating the precedence effect observed that the auditory system tends to focus the core of the localization process on the computation of localization cues carried by the first arriving sound. Doing so relieves the auditory system from dealing with contradictory localization cues in later arriving sounds. A recent study by Dietz et al. (2013) confirmed that human listeners use this approach to deal with dynamic localization cues. In order to provide an explanation for this finding, we first tested several auditory models on the specific task described in Dietz et al. (2013) in order to shortlist possible mechanisms capable of accounting for the early extraction of temporal binaural cues. We found that the best candidates to account for this data are single cell mechanisms, such as adaptation and onset firing, as well as inhibitory population mechanisms. To further understand how each mechanism contributes to the suppression of lagging sounds, we designed more general models capable of demonstrating the principal features of each mechanism. We tested these models thoroughly and found that all mechanisms were able to reproduce the results over a wide range of parameters. This finding suggests that mechanisms responsible for the precedence effect may not be specialized to perform this specific task but instead may be the results of more commonly found neural circuits in the brain. Finally, to facilitate comparing the performance of auditory models on psychoacoustical data, we also designed and implemented an auditory modelling framework capable of addressing many challenges existing in the field of auditory modelling.

Acknowledgements

I would like to express my heartfelt gratitude to the many people who helped me get to this stage in my life. First, I would like to thank my supervisor Dr. Dan Goodman for his excellent guidance, support, trust and availability. Sharing with me his inspiring views on the field shaped me as a scientist in a lasting manner.

Second, I want to thank my labmates for the amazing and heartwarming work environment they created.

In addition, I would like to thank Prof. Mathias Dietz for giving me his valuable feedback on my work.

I would also like to express my profound gratitude to my wife, Vassiki, for always being there for me even in the extraordinary circumstances of living 5252 km away from each other.

Finally, I want to thank my family. My parents, sisters, and grandmother, who encouraged me and had faith in me all the way. Thank you.

Declaration of originality

This thesis is submitted to the Electrical and Electronic Engineering Department of Imperial College London, in fulfilment of the requirements for the degree of Doctor of Philosophy. The contributions contained in this thesis are my own work with the guidance of my supervisor Dr. Dan F.M. Goodman.

Copyright declaration

Contents

List of Tables

List of Figures

Glossary

ϕ_{start} Starting phase. xviii–xx, xxiii, 22, 23, 25, 27, 29, 34, 62

f_c Carrier frequency. 28

f_m Modulation frequency. 21

AFC A Forced Choice toolbox. 87

AM Amplitude Modulated. 15

AMBB Amplitude Modulated Binaural Beats. xviii, xix, xxiv, xxv, 13–15, 21, 22, 27, 41, 42, 53, 59, 61, 63, 67, 71, 73, 79–81

AMT Auditory Modelling Toolbox. 88

ANF Auditory Nerve Fibre. 8, 25, 30, 35

BIPD Best Interaural Phase Difference. xii, 10, 38, 63, 67, 68, 70, 72

BMLD Binaural Masking Level Difference. 18, 87, 88

BSS Blind Source Separation. 16, 17

CASA Computational Auditory Scene Analysis. 17

CD Coincidence Detector. 9, 35, 36

CI Cochlear Implant. 18, 90, 91

DCN Dorsal Coclear Nucleus. 8

ERB Equivalent Bandwidth Rectangle. 22

HRTF Head Related Transfer Function. 5

IC Inferior Colliculus. xvii, 7, 12, 15, 35, 80, 81

ICA Independent Component Analysis. 16, 17

IHC Inner Hair Cell. xix, 25, 26

ILD Interaural Level Difference. 5, 6, 9, 12, 14, 80, 92

IPD Interaural Phase Difference. xi, xviii, xix, 5, 6, 9, 10, 13–15, 21–24, 27–30, 33, 35–39, 41, 42, 49, 51, 55, 58, 59, 61, 63, 65, 67, 68, 73–75, 77, 79–81

ITD Interaural Time Difference. x, 5–7, 9, 10, 12, 14, 18, 19, 22, 39, 80, 82, 87, 88, 92

LIF Leaky Integrate and Fire. 35

LSO Lateral Superior Olive. 9, 30, 92

MGB Medial Geniculate Body. 11

MSO Medial Superior Olive. 9, 12, 15, 35, 80, 81

NCC Normalized Cross Correlation. 22, 28

OHC Outer Hair Cells. 21

PE Precedence Effect. 1

PSTH Peristimulus Time Histogram. xx, 30, 34

rBMF Rate Best Modulation Frequency. 58

rMD Rate Modulation Depth. 47, 49, 57

rMTF Rate Modulation Transfer Function. 47, 56

tBMF Temporal Best Modulation Frequency. 57

tMD Temporal Modulation Depth. 47, 49

tMTF Temporal Modulation Transfer Function. 47, 49

VCN Ventral Coclear Nucleus. 8

WAV Waveform Audio File Format. 84

Chapter 1

Introduction

1.1 Introduction

1.2 Definition of the problem

Sound localization is an important evolutionary advantage for many species. Animals use it to spot nearby predators or prey and more generally to get a better representation of their surroundings (Darwin, 2005). The type of computations required to perform sound localization involves the estimation of time and level differences between the sounds reaching the two ears (Macpherson and Middlebrooks, 2002). The specifics of this process are not yet fully understood in mammals (Grothe et al., 2010) and additionally this task increases in complexity in the presence of reverberations since the auditory system is required to disambiguate the confounding localization cues carried by the reflections in order to accurately assess the position of the sound of interest (Shinn-Cunningham and Kawakyu, 2003). A wide range of species, including humans, are capable of robustly localizing sounds in the presence of reverberations (Hartmann, 1983; Rakerd and Hartmann, 2005). This ability is related to a set of perceptual behaviours referred to as 'the precedence effect' (PE) (Wallach et al., 1949). The precedence effect describes the capability of the auditory system to focus the core of the localization process on the first arriving sound. Doing so unburdens the auditory system from dealing with

confounding localization cues carried by reverberations. The precedence effect has been widely studied over the years in many different fields including animal and human psychophysical experiments, as well as computational modelling (Zurek, 1987; Blauert, 1997; Litovsky et al., 1999; Brown et al., 2015). As a result many models have been proposed to account for it. None of them, however, were capable of explaining the precedence effect in its entirety (Braasch and Blauert, 2003; Brown et al., 2015). This result may suggest that more than one neural mechanism are in fact responsible for this effect. We examine here the performance of a set of commonly found neural mechanisms in the brain on extracting temporal localization cues in reverberative sounds.

1.3 Approach and challenges

The field of auditory modelling offers a wide variety of computational models, describing multiple stages of processing of the auditory pathway (Meddis et al., 2010; Verhulst et al., 2018). These models are useful exploratory tool to tackle many aspects of auditory science. A shortcoming of computational modelling however, is that models are often task dependent and therefore difficult to use outside of the appropriate experimental context (Dietz et al., 2018). Moreover, numerous models rely on many parameters, often tuned to match physiological data (Rothman and Manis, 2003; Bruce et al., 2018). This often makes it difficult to comprehensively grasp the range of behaviours the model can display (O'Leary et al., 2015). We prioritized here the use of general, canonical models of neural mechanisms, relying on fewer parameters and facilitating the full exploration of the parameter space of the model and consequently its full range of behaviours. Taking this approach increases the exploratory power of the models but also limits their generality. This issue is addressed in chapter 7.

1.4 Thesis overview and contributions

The work described in this thesis aims at improving our understanding of the way the human brain computes temporal localization cues in reverberative sounds. To this end, we review in chapter 2 the existing literature on the topic of sound localization, the precedence effect and the cocktail party effect. Following that, in chapter 3, we scout for potential neural mechanisms capable of accounting for the experimental data reported in a psychoacoustical study investigating the extraction of temporal binaural cues in amplitude modulated sounds Dietz et al. (2013). The findings described in chapter 3, allowed us to identify three canonical neural mechanisms that appear to be good candidates to explain the subset of the precedence effect we are interested in. These three mechanisms are : adaptation, onset firing and inhibition. Two of these mechanisms, adaptation and onset firing act at the level of a single cell whereas inhibition takes place in a population of neurons. We examine in chapter 4 and 5, simple, general models of each mechanism and study their behaviour in details through a comprehensive exploration of their parameter space. Chapter 4 focuses on the effect of adaptation and onset firing on the perceived sound source location while chapter 5 investigates the effect of central inhibition. Chapter 4 and 5 are adapted from the preprint study Lestang and Goodman (2019). Later, in chapter 6, we discuss in more details the findings described above. Next, in chapter 7, we detail the design and implementation of a useful framework we developed capable of addressing many challenges currently existing in the field of auditory modelling. Chapter 7 is partly adapted from Dietz et al. (2018). Lastly, in chapter 8, we described potential applications for our results and present possible future directions for our research.

Chapter 2

Background Theory

This chapter contains content adapted from my preprint study (Lestang and Goodman, 2019).

Hearing is an effective tool to scan the surrounding environment (Bregman, 1990). Many informative cues can be inferred from the estimation of the physical properties of the perceived sounds. The perception of the frequency of sounds, for example, facilitates the process of object grouping in the analysis of auditory scenes (Bregman, 1990; Darwin, 2005) and is essential in the processing of speech. Similarly, many species are also capable of inferring the location of a sound from the signals reaching the ears. By estimating the level and time differences of the signals between the two ears, in addition to spectral cues, the brain can estimate the location of incoming sounds (Blauert, 1997).

2.1 The functioning of sound localization in the auditory system

2.1.1 Localization cues

The ability to localize sound derives from the extraction of three main localization cues (Blauert, 1997).

Spectral cues First, while travelling from the source to the ears, sounds undergo a series of transformations that affects the distribution of their spectral components. The shape of the ears, head and torso, are responsible for many of the transformations the sounds is subjected to before reaching the eardrum (Rice et al., 1992; Hammershøi and Møller, 1996). The way the spectrum of the sound is affected by these transformation is a first cue to the location of its source (Middlebrooks, 1992). These cues are referred to as monaural spectral cues and their transfer function, the ratio of the Fourier transform of the sound reaching the ears to the Fourier transform of the original sound, is referred to as the head-related transfer function (HRTF).

Binaural cues The last two cues are computed by comparing the signals reaching the left and right ears. Because of the geometry of the head and the position of the ears, the signal arriving at each ear is likely to differ. As a result, depending on the location of the source of the sound, the signals arriving at one ear might be delayed in comparison to the signal arriving at the other ear. A sound originating from the left of a listener would reach the left ear before the right ear. Additionally, the sound would likely be louder at the left ear than at the right ear. Because of these properties, these two cues are referred to as interaural time differences (ITDs) and interaural level differences (ILDs). In pure tones, because of the peridiocity of the signals, ITDs can be interpreted as interaural phase differences (IPDs). Phase and time differences, however, can be ambiguous at certain frequencies since the phase difference between the two left and right channels can be interpreted in either direction (figure 2.3.1) resulting in mapping

the cues to two different locations. Fortunately, for low-frequency sounds, this ambiguity is resolved as the maximum IPD perceivable for the animal leads to an IPD that is less than half a stimulus cycle. The range of IPD values perceivable by an animal is referred to as the physiological range of the species.

2.1.2 The duplex theory

The contribution of each cue to the estimation of the location of the sounds largely depends on the spectral composition of the sound and the location of the source (Macpherson and Middlebrooks, 2002). First, since low-frequency sounds are not remarkably attenuated by the head, ILDs tend to be too small at low-frequencies to be accurately detected and used to localize sounds. This results in ITDs being mainly used to localize low-frequency sounds (Rayleigh, 1875). High frequency sounds, on the other hand, impair the ability of the auditory nerve fibers to fire in phase with the frequency of the sound (Köppl, 1997; Verschooten et al., 2019). This makes the ITD information conveyed by the fine structure of the sound unusable at high frequencies. However, envelope ITDs can still detected at high modulation frequencies (Bernstein and Trahiotis, 1994, 2002). As a result, listeners generally rely on interaural level differences to localize high-frequency sounds. This arrangement is often referred to as the duplex theory (Rayleigh, 1875; Macpherson and Middlebrooks, 2002). Lastly, monaural spectral cues are useful spatial cues as they allow monaural sound localization (Wightman and Kistler, 1997) and permit listeners to estimate the elevation of sounds under certain conditions (Roffler and Butler, 1968).

2.2 Physiology and neuroanatomy of ITD computation in sound localization

The auditory pathway is composed of a set of functional structures that provides efficient mechanisms to extract localization cues through a multi-stage processing of the incoming sound.

The main structure of the auditory systems are the inner, middle and outer ear, several regions of the brainstem, the thalamus and the auditory cortex (figure 2.1)

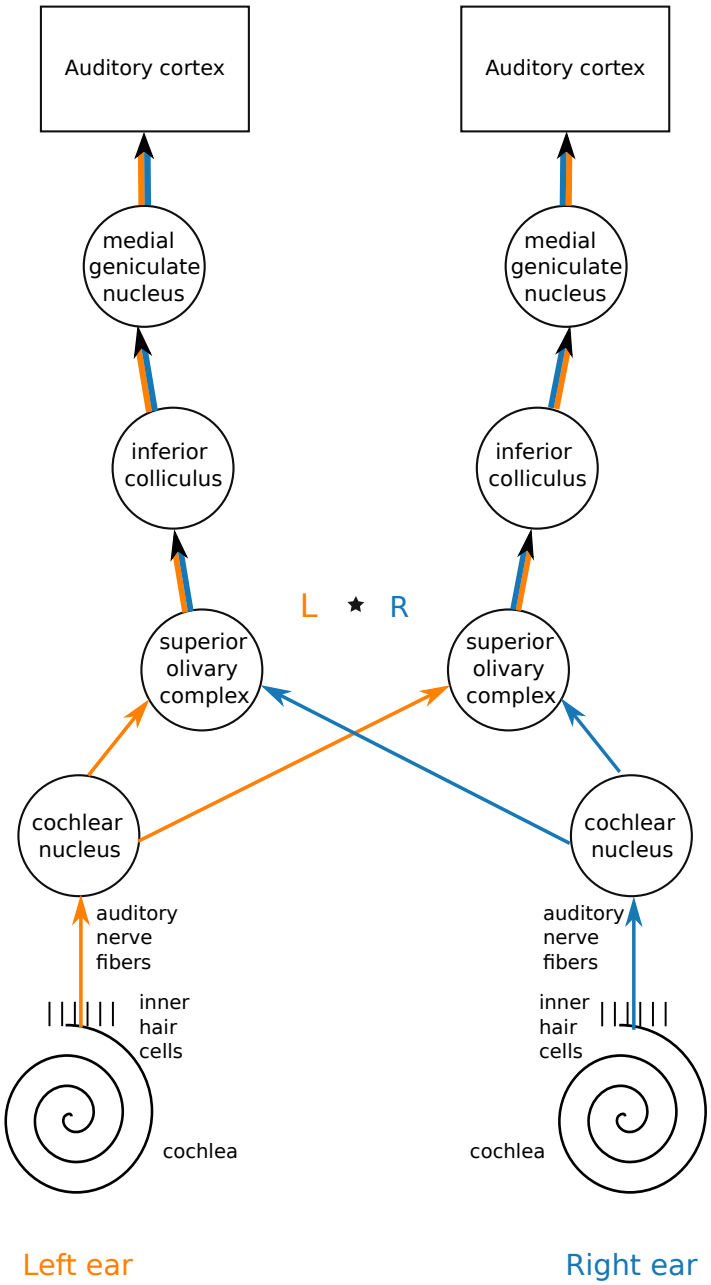

Figure 2.1: Simplified schematic of the auditory pathway. From the cochlea to the cochlear nucleus, only monaural processing is taking place. At the level of the superior olive, signals coming from the left and right ear are compared through a coincidence detection process similar to cross-correlation (Colburn, 1973). The signals, next, converge towards the inferior colliculus (IC). After the IC the signals reach a region of the thalamus called the medial geniculate nucleus and finally converge toward the auditory cortex.

2.2.1 The periphery

The sounds arriving at the ears are first transformed by the ear itself, the pinna, before transitioning to the ear canal. These two structures constitute the outer ears. At the end of the ear canal, the sound causes the vibration of a membrane, the eardrum. The vibration of the eardrum causes the displacement of three ossicles (the malleus, incus and stapes) which work together in focusing the energy of the sound on a smaller spot located on the stapes. The stapes, in response, percusses the oval window of the cochlea causing the fluid inside the basilar membrane of the cochlea to move. The displacement of this fluid, induces the movement of hair cells which depolarize the auditory nerve fibers (ANFs) causing them to fire. The spatial arrangement of the basilar membrane leads to a frequency decomposition of the signal. The basilar membrane exhibits a gradient of preferred frequency from its base, where the preferred frequency is around 20000 Hz, to the apex where the best frequency is around 200 Hz. This results in auditory nerve fibers responding best to specific frequencies. At low-frequency the spikes triggered by the ANFs are in phase with the sound frequencies. This property, referred to as phase-locking, is not true at high frequencies. Due to this frequency decomposition taking place early on in the auditory system, the rest of the auditory pathway is mainly organized in a manner that ensures that similar frequencies are processed in neighboring regions. This organization, also known as tonotopy, is kept throughout the auditory pathway, starting in the cochlea and up to midbrain structures (Kandler et al., 2009), as well as some parts of the auditory cortex (Saenz and Langers, 2014). The next step of processing is located in the cochlear nucleus, composed of the dorsal and ventral cochlear nucleus (DCN and VCN). Both DCN and VCN are known to contain a wide variety of cells. Chopper cells, located in the VCN, have been shown to demonstrate very precise firing in response to periodic stimuli. This property makes the cochlear nucleus particularly well suited to encode temporal envelope cues (Lorenzi et al., 1995; Carney et al., 2015). Because the cochlear nucleus is located immediately after the ANFs, it is also believed to be the main information relay from the ANFs to brain regions located higher in the auditory pathway.

2.2.2 The midbrain and the computation of interaural time differences (ITDs)

Delay lines arrangement After the cochlear nucleus, the signals coming from both ears are processed together to compute the binaural cues ITD and ILD. The medial superior olive (MSO) located in the midbrain is believed to mainly deal with the computation of interaural time differences while the lateral superior olive LSO is presumed to be the center of ILD computations. The MSO and LSO are the first structures to have as inputs signals coming from both sides of the head. It has been proposed that the computation of the ITD is performed in the MSO via coincidence detection. Coincidence detector (CD) neurons are cells that respond to simultaneously arriving spikes from both ears (Jeffress, 1948) (figure 2.2). This process was first suggested by Jeffress who designed a model relying on delay lines and coincidence detection to compute ITDs (Jeffress, 1948). According to the Jeffress model, neurons code for the azimuth of a sound via their firing rate. Neurons displaying the highest firing rate code for the delay best represented in the sound, and therefore for the azimuthal location of the source. The neurons overall respond best to simultaneously arriving inputs from both ears. In this model, MSO neurons on each side of the head show a gradient in axonal delays. On each side of the head, the gradients are directed towards opposite direction. This arrangement results in an array of coincidence detector neurons, each coding for a specific delay (figure 2.2). This mechanism has been shown to be similar to a cross-correlation operation between the left and right channels (Colburn, 1973).

Limits of the delay lines arrangement Although delay lines were observed in the barn owl (Carr and Konishi, 1988; Konishi, 2003), the existence of delay lines arrangements in mammals has not been quite established (Grothe et al., 2010). First, (Brand et al., 2002) showed that at the level of the MSO, precise inhibition is critical in the detection of IPDs in gerbils. Blocking this inhibitory input, shifted the tuning of the cells towards a 0° IPD. Because in theory the Jeffress model only requires precise excitatory inputs, the necessity of the presence of inhibition to produce various ITD tuning is surprising. In addition, it has

been observed that a large portion of neurons in the inferior colliculus of guinea pigs exhibit large ITDs tuning outside of the physiological range of the species (McAlpine et al., 2001). In contrast, tuning to ITDs within the physiological range of the species were only observed in the slope of the tuning curve of the cells, suggesting a slope encoding of the IPDs in mammals. Additionally, the authors observed that the distribution of the best interaural time differences BIPDs is largely dependent of the frequency channel. These findings question the role of such cells in a delay lines arrangement and the existence of peak ITD tuning in mammals (McAlpine and Grothe, 2003). In view of these results, estimating the ITDs from the most active cell alone does not appear to be the most optimal strategy for mammals. Instead it has been proposed that the decision on the location of the sound is estimated at the population level. In response to that, several population decoder have been suggested. The hemispheric decoder (Stecker et al., 2005; Grothe et al., 2010; Lesica et al., 2010) hypothesizes that ITDs are encoded by comparing the activity of the two brain hemispheres. Another recently proposed mechanism conjectures the existence of a firing rate pattern-matching operation taking place in populations of ITD sensitive neurons (Goodman et al., 2013).

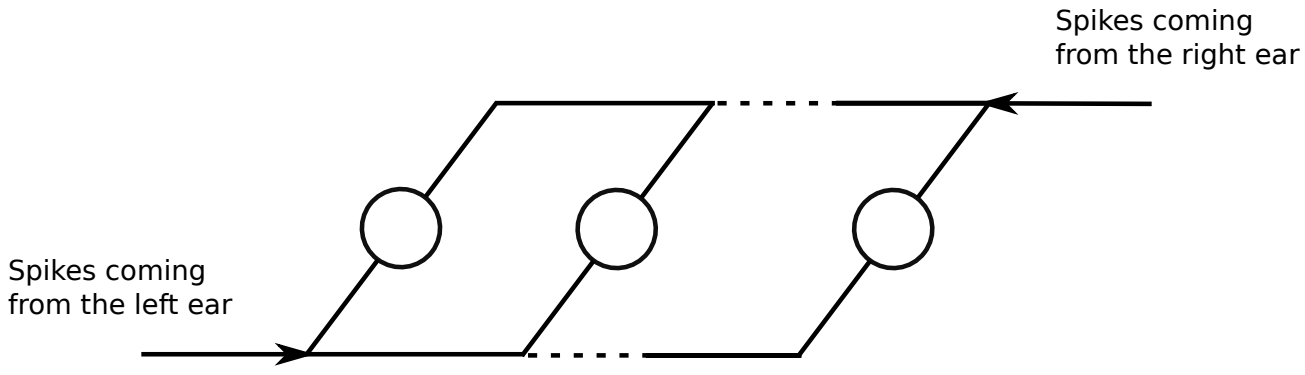

Figure 2.2: Jeffress model, spikes coming from the left and right ears converge on coincidence detector neurons (circles). The length of the axons from both sides varies gradually in opposite directions ensuring that each neuron responds best to a specific delay. This delay maps the azimuthal location of the sound

2.2.3 Higher structures

Higher up in the auditory pathway, the medial geniculate body (MGB) has been shown to respond to localization cues in cats (Ivarsson et al., 1988; Clarey et al., 1995). However, in rats, the activation of the MGB has been observed to not be critical to perform sound localization tasks (Kelly and Judge, 1985). Some parts of the mammalian auditory cortex are known to play a role in sound localization. This has been shown in the past by removing or impairing the auditory cortex of animals and testing their ability to localize sounds (Whitfield et al., 1972; Heffner, 1997; Malhotra and Lomber, 2007). Interestingly, one of the main finding in this field of study is the absence of neural fields directly mapping the location of the sounds to specific neurons (Middlebrooks and Pettigrew, 1981) reinforcing the idea that sound location is encoded by population of neurons instead of single cells (Stecker et al., 2005).

2.3 Precedence effect

2.3.1 Psychophysical findings related to the precedence effect

The precedence effect, or law of the first wave, is a set of perceptual phenomena which describe the perception of the location of sounds when multiple sounds reach the listener's ears in quick succession (Wallach et al., 1949). It has been reported that the delay between the leading and the lagging sound greatly influences the observed behaviour (Litovsky et al., 1999; Brown et al., 2015). Generally speaking, three main behaviours have been observed from varying the lead-lag delays. When the delay between the two sounds is very short (below 1 ms), the listener hears a fused sound whose location is perceived between the two sound sources. This phenomenon is referred to as *summing localization*. When the delay between the two sounds is slightly longer (around 1 ms), the listener tends to perceive a single sound whose location is positioned near the source of the leading sound. This behaviour has been referred to as *localization dominance* (Wallach et al., 1949; Litovsky et al., 1999). Finally, for delays large enough, the two sounds can be discriminated and the locations of both sounds can be estimated by the listener. The

smallest delay allowing the listener to segregate the two sounds is referred to as the echo threshold. Because the precedence effect directly affects the perception of the location of the incoming sounds, understanding how the computation of localization cues are affected by this phenomenon is pertinent. By measuring the discrimination threshold (echo threshold) between sounds lateralized with either ITD or ILD, several studies showed that longer delays are needed to reach the echo threshold with ITD lateralized sounds rather than with ILD lateralized sounds (Zurek, 1980; Krumbholz and Nobbe, 2002; Brown and Stecker, 2013), suggesting that ITD lateralized sounds are more affected by localization dominance than sounds lateralized with ILD only. Understanding how the computation of ITDs is performed when multiple incoming sounds are reaching the ears is therefore relevant to explain the precedence effect in more details.

2.3.2 Mechanisms likely to be involved in the precedence effect

Although well studied (Zurek, 1987; Blauert, 1997; Litovsky et al., 1999; Brown et al., 2015), the neural mechanisms involved in 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. Neural evidence of echo suppression have been shown to exist in many different parts of the auditory pathway, starting as early as the periphery and the auditory nerve fibers (Fitzpatrick et al., 1999) up to the medial superior olive and inferior colliculus (Dietz et al., 2014; Yin, 1994). 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 confounding localization cues by only responding to the onset of sounds. On the other hand, central models of the auditory system including inhibitory circuits in the medial superior olive (MSO) or inferior colliculus (IC) were

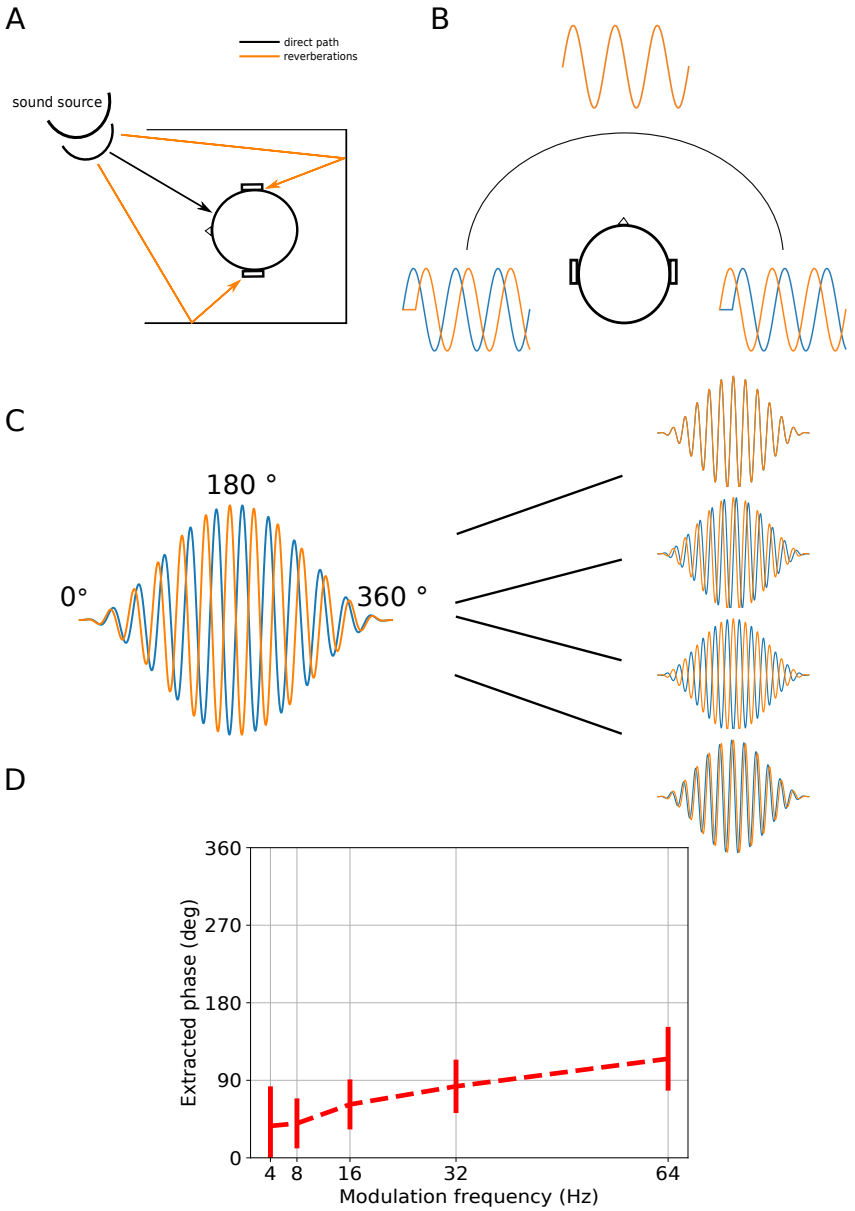

Figure 2.3: (A) Despite reverberations carrying misleading localization cues, listeners can locate the source of a sound. (B) Relationship between interaural phase differences and lateralization. When the sound is located in front of the listeners the sound arrives at the ears at the same time, inducing no IPD. However, when the sound arrives at the ears from the sides, an IPD is induced as the sounds arrive at the ears at different times. (C) Left: Example of amplitude modulated binaural beats. Here the carrier frequency was chosen equal to 10 Hz and the modulation frequency was equal to 4 Hz. The IPD starts at 0° at the start of the envelope and ends at 360° at the end of the envelope. Right: The perceived location of the AMBB sound is compared to the one produced by AM sounds carrying static IPD. The static IPD AM sound maximizing the similarity is chosen to extract the phase at which the IPD is perceived by the listener. (D) Results of the psychoacoustics task from (Dietz et al., 2013). All IPDs extracted are below 180 ° (the peak of the envelope) and the IPD extracted increases with modulation frequency. Error bars represent the circular mean of the standard deviation across subjects

also able to account for some aspects of the precedence effect. A large part of the models proposed rely on the transient inhibition of the lagging sound (Lindemann, 1986a,b; Breebaart et al., 2001; Xia et al., 2010) in inhibitory networks located in the midbrain. Lindemann's model (Lindemann, 1986a,b) extends the Jeffress model to the computation of ILD by adding contralateral inhibition and monaural detectors. A different approach was taken in Xia et al. (2010), the authors designed a detailed model of the auditory pathway including inhibitory and excitatory connections between the two hemispheres. This arrangement causes the suppression of the lagging sounds to directly depend on the location of the sounds and the lead lag delay. Overall, peripheral models were found to account well for precedence effect related behaviours with short sounds while inhibition based models were shown to produce better results with longer sounds (Braasch and Blauert, 2003; Brown et al., 2015). Despite occurring at different stages of the auditory pathway, a large portion of the mechanisms listed above mainly consists in a combination of canonical neural circuits such as adaptation, onset firing and inhibition.

2.3.3 Early extraction of temporal cues in low-frequency amplitude modulated signals

In order 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 (IPD) 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 the time-course of a single envelope cycle (figure 2.3.1C), while ensuring that no ILDs are available. The authors used a carrier frequency of 500 Hz which is within the range of frequencies allowing reliable ITD extraction. Additionally, the modulation frequency ranged from 4 Hz to 64 Hz. By asking participants to match the perceived location produced by the AMBB sounds to static IPD pointers (figure 2.3.1C), 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 energy of the

sound is maximal). The perceived IPD was also shown to increase with modulation frequency (figure 2.3.1D). More recently, the authors also recorded the responses of MSO and IC cells in anesthetized gerbils and guinea pigs to the AMBB stimulus (Dietz et al., 2014). They found that a large portion of cells responded best to IPDs located in the rising segment of sounds while the remaining neurons responded best to the IPD located at the peak. These results are in line with previous findings by Devore et al. (2009) who observed that, in reverberant environments, cells in the IC of anesthetized cats reached their maximal sensitivity to directional cues in the onset of sounds. Later, Hu et al. (2017) conducted a similar experiment to Dietz et al. (2013) and uncovered that IPD extraction in the rising slope depends on the value of the carrier frequency. In AM sounds with a carrier frequency equal to 200 Hz and modulation frequencies equal to 8 Hz or 20 Hz, the subjects mostly reported IPD located at the peak of the envelope. However, for 600 Hz stimuli, subjects prioritized the rising portion of the envelope.

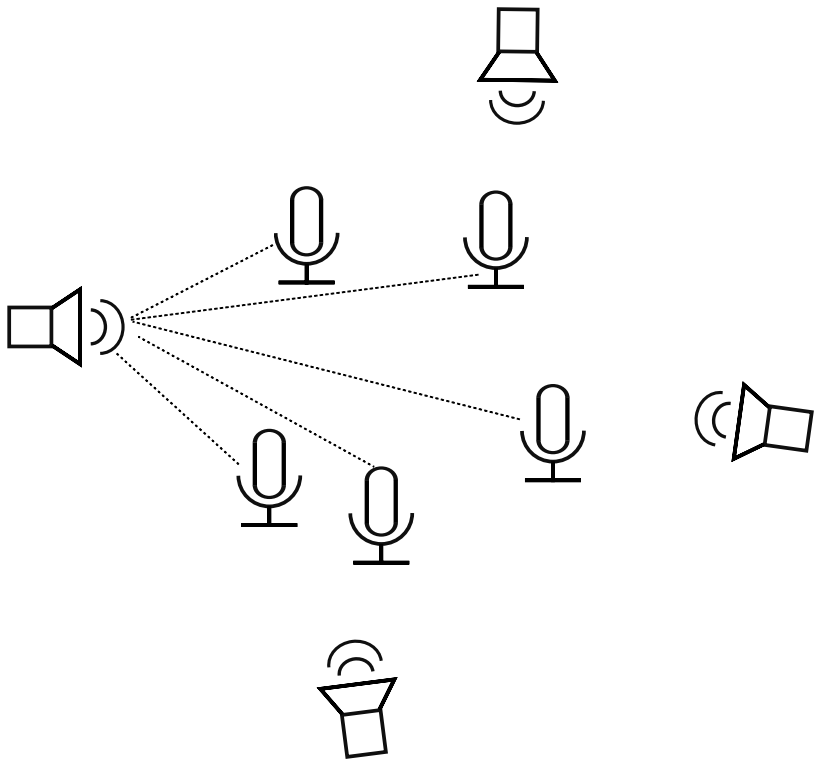

Figure 2.4: Cocktail effect and blind source separation. The task is to segregate the n signals coming from the n speakers ($n = 4$ in the figure) from the m mixtures received at each of the m microphones ($m = 5$ in the figure).

2.4 Relevance of understanding the precedence effect

The precedence effect is related to another interesting perceptual behavior: the ability to focus our attention on a particular sound. This ability is known as the cocktail party effect (Cherry, 1953b). In addition, a better understanding of the precedence effect is likely to be beneficial to the improvement of hearing aids.

2.4.1 Cocktail party effect

A remarkable human behaviour is the ability to selectively focus attention on an auditory stimulus of interest in the presence of competing stimuli. In order to understand the challenges of the cocktail party effect, we will first focus on formalizing the problem mathematically and review the current signal processing methods currently used to solve it. Following that, we will show how accurately localizing sounds helps with demixing signals in cocktail party situations.

Blind Source Separation In this section, we are examining more closely how the cocktail party effect is usually solved in the field of signal processing. The aim of doing that is to decide whether the brain is likely to use existing methods (such as Independent Component Analysis ICA described below) to solve this problem. In the field of signal processing, the problem of separating a mixture of signals, with no prior knowledge on the nature of the signals, is referred to as blind source separation (BSS). Considering n sources and m sensors, the aim of BSS is to reconstruct each original signal from the m sensors recordings. An illustration of the problem can be seen in figure 2.3.3. From the description of the problem, several questions arise, including:

- How does the nature of the sources influence the demixing process ?
- Is there any critical number of sensors needed to accurately demix the signals ?

One commonly used algorithm for solving BSS is the independent component analysis algorithm (ICA) which we will briefly discuss here. Performing ICA generally requires several assumptions:

- The source signals should be linearly mixed and independent from one another.
- At most one source can follow a gaussian distribution.
- Finally, m, the number of sensors, should be at least equal to n the number of sources.

The general principle of ICA is to extract the original signals by finding which linear combination of potential sources minimizes the gaussianity of each source (Hyvärinen et al., 2001). The idea is that gaussian distributions maximize entropy amongst all distributions with unitary variance. Because of the non-gaussianity assumption, it becomes possible to find the set of sources that minimizes gaussianity. Although ICA has been shown to work well (Hyvärinen et al., 2001), its functioning relies on setting up strong assumptions which in some situations can not be met ($m > n$ for instance). It is not clear whether the brain performs ICA inspired computations. Although it has been shown that spiking neural networks implementing ICA can be designed (Savin et al., 2010; Clopath et al., 2008), the fact that the brain only uses the signals from the two ears to extract the speech channel of interest makes it unlikely. Because solving the BSS problem has been shown to be a difficult task, understanding in more details how the brain solves the cocktail party effect could lead to an useful insight on how to solve the more general problem of source separation.

Grouping cues in cocktail party situations The field of computational auditory scene analysis (CASA) breaks down the extraction of a target-speech in a cocktail party environment into two separate steps (Bregman, 1990; Bronkhorst, 2015). The first one consists in segregating in real-time the different incoming signals. The second step is the task of streaming which consists in joining the different extracted speech segments together. Segregating speech channels involves computing informative features to cluster speech segments based on their origin. For instance, Darwin et al. (2003), showed that speech segregation is easier to perform when the fundamental frequency of the speech produced by the speakers differ. Additionally, it has been shown that listeners are better at extracting a target stimulus from a masking one when the sounds are spatially separated (Licklider, 1948; Freyman et al., 1999; Hawley et al., 2004). This phenomenon is known as spatial release from masking and appears to be an important

mechanism used by the brain to solve the cocktail party problem. An important finding related to spatial unmasking is the binaural masking level difference (BMLD). In a task where a listener is asked to detect a target signal from corrupting noise, the detection of the signal improves when the phase of the signal or its level are different to the masker's. The phase difference produces the perception that the signal and the noise are originating from two different locations and contribute to improving the detection threshold (Hirsh, 1948; Licklider, 1948). Similar strategies are also used in the the field of signal processing. They are referred to as beamforming techniques (Johnson and Dudgeon, 1993) and consist in the design of a spatial filter which enhances sounds coming from a specific direction through constructive interference and discards sounds from other directions via destructive interferences.

2.4.2 Cochlear implant technology

Normal-hearing listeners are able to accurately localizing sounds in the presence of mild reverberations (Hartmann, 1983; Rakerd and Hartmann, 2005). However, this process is more problematic for hearing impaired listeners (Lorenzi et al., 1999). The overall performance in localization tasks has been shown to be much lower in bilateral cochlear implant (CI) users compared to normal-hearing listeners (Grantham et al., 2008). These unsatisfactory performances can be attributed to the fact that bilateral cochlear implant listeners struggle to extract interaural time differences to locate sounds and tend to mainly rely on interaural level differences (Grantham et al., 2008). Because accurate sound localization in reverberative environments is believed to depend on a satisfactory sensitivity to ITD, it is not surprising that bilateral CI listeners perform poorly in localizing sounds in the presence of reverberations (Brown et al., 2015). Additionally, improvements in ITD sensitivity in cochlear implants technology alone do not guarantee the reproducibility of the perceptual ITD weighting observed in normal hearing listeners (Dietz et al., 2013; Hu et al., 2017). In fact, Hu et al. (2017) showed that bilateral cochlear implant listeners rely on ITD information located at the peak of the envelope of the stimulus, even at carrier frequencies where normal-hearing listener prioritize the temporal information located in the rising part of the envelope. Understanding the neural mechanisms

used by the brain to handle the suppression of lagging sounds is therefore likely to provide auditory researchers valuable insights to improve the design of cochlear implants.

2.5 Summary

Sound localization relies on the extraction of three different cues, namely spectral cues, interaural level differences and time differences (section 2.1.1). The computation of one of them, the interaural time difference (ITD), is not yet understood in the presence of reverberations. The precedence effect, a perceptual phenomenon related to this problem, suggests that the auditory system prioritizes the extraction of the ITD carried by the first arriving signal at the ears. Several mechanisms have been proposed over the years to account for this 'law of the first wave'. Most of them rely on canonical neural dynamics such as adaptation, onset behaviour and inhibition. A psychoacoustical study published in 2013 (Dietz et al., 2013), confirmed the finding that normal-hearing listeners rely on ITDs located in the rising portion of amplitude modulated sounds. An important part of the work described here attempts to explain the results uncovered by the authors. A better understanding of the mechanisms underlying the precedence effect has the potential to produce valuable insights on many other problems, including the design of cochlear implants, solving the cocktail party problem or more generally improving signal demixing techniques in signal processing.

Chapter 3

Search for suitable mechanisms

I have presented work based on this chapter during a poster session at the 2017 Acoustical Society of America meeting in Boston. (Lestang and Goodman, 2017)

Auditory research has uncovered over the years many different mechanisms, occurring at different levels of the auditory pathway, capable of accounting for different perceptual and neurophysiological behaviours. Computational models are designed to emulate these mechanisms. In this chapter, we examine the performance of several models in reproducing the data described in Dietz et al. (2013) and Hu et al. (2017) and scout out for neural mechanisms capable of extracting temporal binaural cues in sound onsets. Mainly, these models can be divided into two categories: peripheral and central. Peripheral models tend to describe auditory mechanisms occurring in early parts of the auditory system, from the inner ear up to the cochlear nucleus. On the other hand, central models refer to stages of processing located higher up in the auditory pathway, from the brainstem up to the inferior colliculus and auditory cortex.

3.1 Peripheral models

The auditory pathway is composed of multiple stages of processing. In this section, we examine the effect of each step of processing that occurs at the periphery and analyse their effect on

emphasizing the onset portion of AMBB sounds.

3.1.1 Outer hair cell compression

Compression is an important feature of the auditory system as it allows the brain to process sounds within a wide range of levels. Compression appears to mainly occur in the periphery of the auditory system (Bacon et al., 2004), and in mammals, it is believed to largely take place in the cochlea via the displacement of the outer hair cells (OHCs) (Rhode, 1971; Ruggero et al., 1990). The OHCs amplify low-intensity sounds more than loud sounds resulting in a non-linear compression of the signals over different sound levels. We investigate in this section the effect of a simple model of OHC compression on the IPD extraction in AMBB stimuli.

Methods The models described in this chapter were implemented in Python. Simulations were performed using the Brian Hears package (Fontaine et al., 2011).

As in Dietz et al. (2013), AMBB stimuli with modulation frequencies (f_m) ranging from 4 Hz to 64 Hz and with a carrier frequency of 500 Hz were used. The sounds followed the equations:

$$E(t) = \tfrac{1}{2}(1 - \cos(2\pi f_m t)) \tag{3.1}$$

$$S_R(t) = \sin(2\pi(f_c + f_m/2)t)\ E(t) \tag{3.2}$$

$$S_L(t) = \sin(2\pi(f_c - f_m/2)t)\ E(t). \tag{3.3}$$

$E(t$ refers to the envelope of the sounds while $S_R(t)$ and $S_L(t)$ refer to the carrier channel of the stimulus. Non-linear compression was modelled using a power-law function:

$$output = ([input]^+)^\rho \tag{3.4}$$

In equation 3.4,the operator $[]^+$ refers to the process of half-wave rectification which consists in passing the positive part of the signal and blocking the negative part.

We tested 3 different power-law coefficients ranging from 1 to $\frac{1}{10}$. The sound duration was

set equal to 1 second (f_m cycles) followed by 100 ms of silence. The frequency decomposition performed by the basilar membrane was modelled using the approximate gammatone filterbank (Hohmann, 2002). Sounds were gammatone filtered around a single frequency band centered around 500 Hz (mean carrier frequency over the two channels). This is reasonable since the spectrum of the sound mainly exhibits a peak around $500Hz \pm \frac{fm}{2}$, and additionally, due to the fact that the equivalent rectangular bandwidth ERB at 500 Hz is equal to 143 Hz (which is higher than the difference of the carrier frequencies of the AMBB stimulus at all modulation frequencies). We applied 125 different delays to the right channel of the sound (figure 3.3A). The purpose of doing that is to generate ITD tuning in coincidence detector neurons and allow the spiking activity of the network to track the dynamic IPD present in the signal (figure 3.7). Because the IPD increases linearly in time following $\phi(t) = 2\pi f_m t$, we obtained the corresponding ITD using $ITD = \frac{IPD}{2\pi f_c}$ which gives us a value of $\frac{1}{f_c}$ at $\phi(t) = 2\pi$. Subsequently, the computation of the interaural phase difference was performed by running a normalized cross-correlation (NCC) operation (Bernstein and Trahiotis, 1996) between the left channel and the delayed copies of the right channel. Normalized cross-correlation is performed in the following way:

$$NCC = \frac{\sum L(t)R(t)}{\sqrt{\sum L(t)^2}\sqrt{\sum R(t)^2}} \tag{3.5}$$

L and R respectively refer to the left and right channels. The NCC operation was performed on the entire duration of the stimulus (1.1 second) and for each pair of left and right signals.

We extracted the IPD maximizing the NCC coefficient. Overall 8 different starting IPDs (IPD value at $t = 0$) were chosen, ranging from 0° to 315°. Starting IPDs will be referred to in the rest of the text as ϕ_{start}. A schematic of the model can be seen figure 3.3A. The curves shown in figure 3.1B are the results of the circular average of the IPDs over the 8 different ϕ_{start}. The circular mean is used for circular quantities and is defined as:

$$\theta = \arctan2(\sum_{i=1}^{N} \sin(\alpha_i), \sum_{j=1}^{N} \cos(\alpha_j)) \tag{3.6}$$

In equations 3.6, α_i and α_j are the angles extracted by the models after each simulation. $arctan2$ computes the arctangent of the ratio of its two arguments. N is the total number of repetitions. θ is the circular average of the different angles.

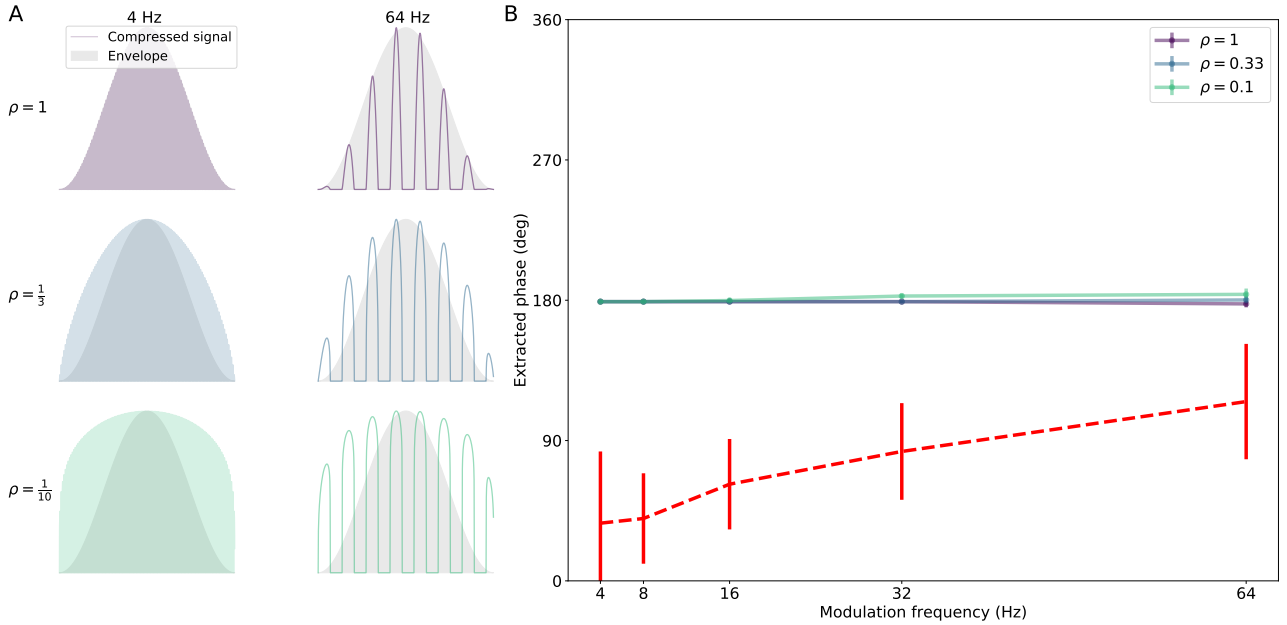

Figure 3.1: (A) Shape of the stimulus after compression when the modulation frequency is equal to 4 Hz (left column) and 64 Hz (right column) and when ϕ_{start}=0. Each row shows the shape of the signal for different values of ρ: $1, \frac{1}{3}, \frac{1}{10}$. (B) IPD extracted when the sound is only subjected to compression with different ρ values.The values displayed are the circular mean of the IPD extracted over the 8 different ϕ_{start}. Error bars of the model are the circular standard deviation over ϕ_{start}

Results The main effect of the power-law compression is to flatten the envelope of the stimulus (figure 3.1A). Smaller power-law coefficients led to flatter envelopes. However, power-law compression alone did not introduce any emphasis on the onset of the sound which led to an extraction of the IPD at the peak of the envelope (180°) (figure 3.1B). Consequently, power-law compression alone can not account for the early extraction of the interaural phase difference in the envelope of the stimulus. This result is not unexpected as power-law compression, is an instantaneous monotonic function of the stimulus. Mechanisms responsible for onset extraction of the IPD therefore must be the product of more dynamic and adaptive processes located at higher stages of the auditory pathway.

3.1.2 Inner hair cell adaptation

Since compression did not account for the emphasizing of the onset of sounds, we examine in this section the effect of neural adaptation occurring at the synapse between the inner hair cells and the auditory nerve fibers on the results.

Adaptation loops

The adaptation loops described in Dau et al. (1996a,b) are also known to implement monaural neural adaptation. Adaptation loops consist in a series of feedback loops composed of capacitors with different time-constants. The input signal is recursively divided by the output of RC-lowpass filters producing a non-linear adaptation (figure 3.2C))

Methods As in section 3.1.1, the sounds were gammatone filtered (Hohmann, 2002) around the carrier frequency and half-wave rectified. Similarly to section 3.1.1, 125 delayed copies of the right channel were created and the IPD extracted maximized the NCC coefficients. The model was run with 8 different values of the starting phase ranging from 0° to 315°.The code we implemented to model the adaptation loops can be seen in appendix 8.3. We also tested the models using random time constants. The five time constants were randomly drawn from five different intervals to ensure a logarithmic distribution of the time-constants. The five time constants were drawn respectively from the intervals [0.001,0.01],[0.005,0.05],[0.01,0.1],[0.05,0.5],[0.1,1] second. They were subsequently sorted in ascending order.

Results At both 200 Hz and 500 Hz, the adaptation loops produced an emphasis of the onset of the signal at low modulation frequencies (figure 3.2A). In addition, the results mainly reproduced the pattern of increasing IPD extracted with increasing modulation frequency (except at $f_m = 64Hz$ at 200 Hz carrier frequency). This increase in IPD extracted is mainly due to a stronger adaptation at low modulation frequencies (figure 3.2A). At 500 Hz, the IPDs extracted were mostly later than those in the data from Dietz et al. (2013) at all modulation frequencies. The best fit of the data produced a maximum absolute error of 85°, which does not constitute

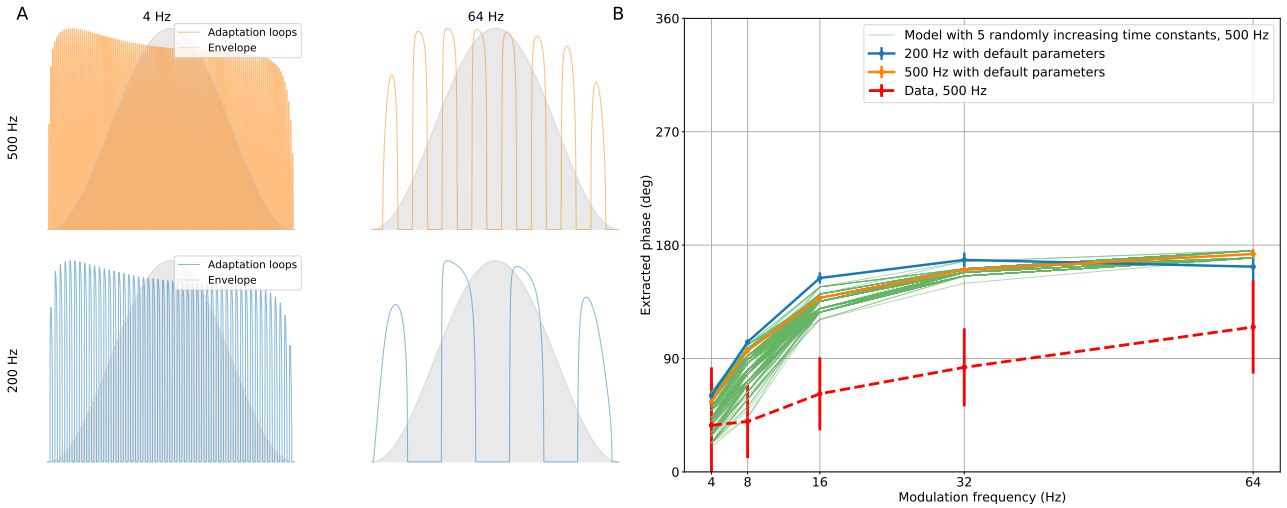

Figure 3.2: (A) Comparison between the envelope of the AMBB stimulus (gray) and the stimulus after half-wave rectification, compression and Dau adaptation (blue and orange) at 4 Hz and 64 Hz. The first row shows the shape of the signal with a carrier frequency of 500 Hz and the second row shows the signals with a carrier frequency of 200 Hz. (B) IPD extracted by the model (with default parameter values) and circularly averaged over the 8 ϕ_{start} and 200 additional phase/f_m curves were plotted with different values for the 5 time constants (green) for a carrier frequency of 500 Hz and a ϕ_{start} of 0°. Error bars of the model are circular standard deviations over ϕ_{start}.

a satisfactory fit of the data. Lowering the carrier frequency from 500 Hz to 200 Hz raised the value of the phase extracted at most modulation frequencies (figure 3.2). This result is qualitatively in line with the results described in (Hu et al., 2017). Since varying the starting phase did not lead to strong variations of the IPD extracted by the model (figure 3.2B), the rest of the results were generated with a single value of ϕ_{start} equal to 0°. The model was subsequently ran 200 times at 500 Hz with random increasing values of the five time constant (between 0.001 second and 0.5 second) which did not lead to large differences in the fit of the data (figure 3.2B).

Meddis adaptation

Many models of inner hair cell adaptation have been proposed in the auditory literature (Zhang and Carney, 2005). A well-known instance is the Meddis model (Meddis, 1986, 1988). It implements the dynamics of neurotransmitters at the synapse between the inner hair cells and the auditory nerve fibres (IHC/ANF synapses). The model can be summarized using three

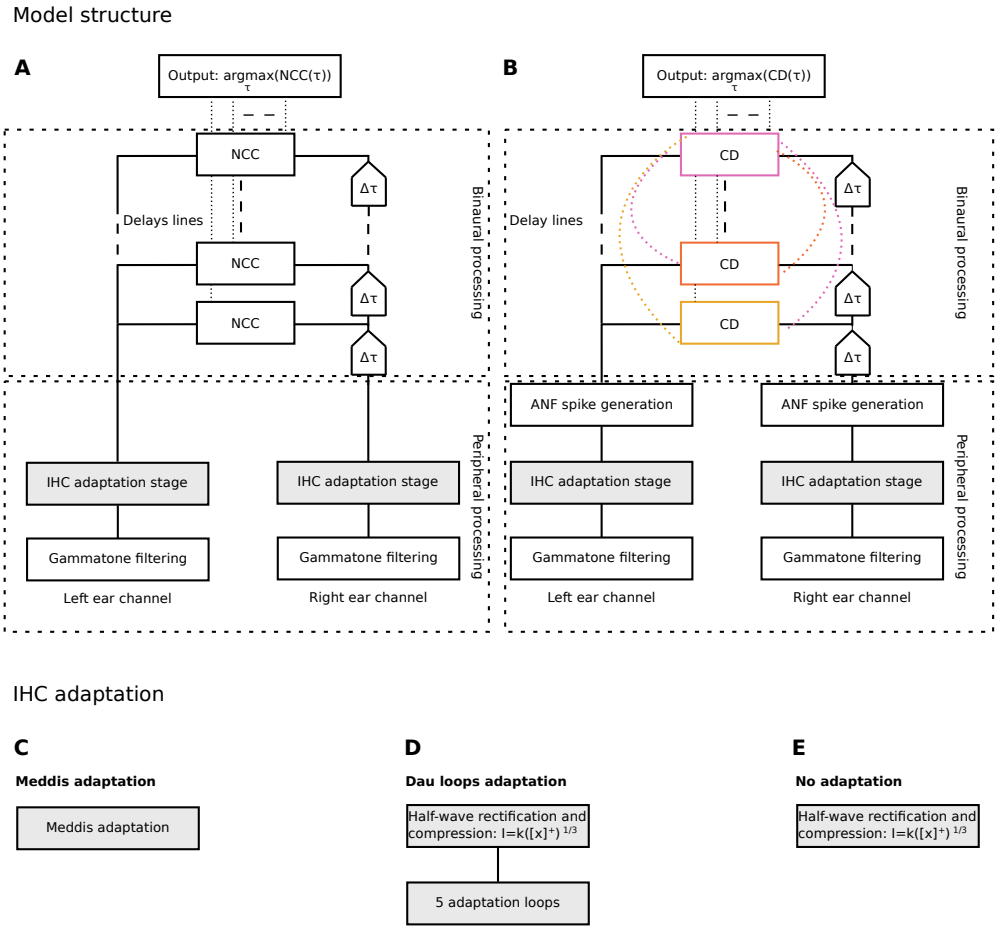

Figure 3.3: (A) Model architecture used in sections 3.1.1,3.1.2. After the peripheral stage, the signals from the right channel are delayed and a normalized cross-correlation (NCC) operation is performed on each channel pair. The pair maximizing the NCC coefficient codes for the extracted phase. (B) Model architecture used in section 3.2. After peripheral processing, the delayed copies of the signals converge on a delay-lines arrangement of coincidence detectors. The coincidence detector that fires the most codes for the extracted phase. (C)-(E) Specifics of the peripheral processing and IHC adaptation stage of processing for each model.

differential equations (equations 3.7, 3.8, 3.9) representing the dynamics of three transmitter reservoirs (figure 3.4E). The first reservoir is the transmitter factory which replenishes the free transmitters pool (instantaneous amount of transmitter in the free reservoir is equal to q) at the rate $y(1-q)$. The second pool reprocesses the transmitters that are not being used at the synapse. A quantity $xwdt$ of neurotransmitters (w being the instantaneous amount of

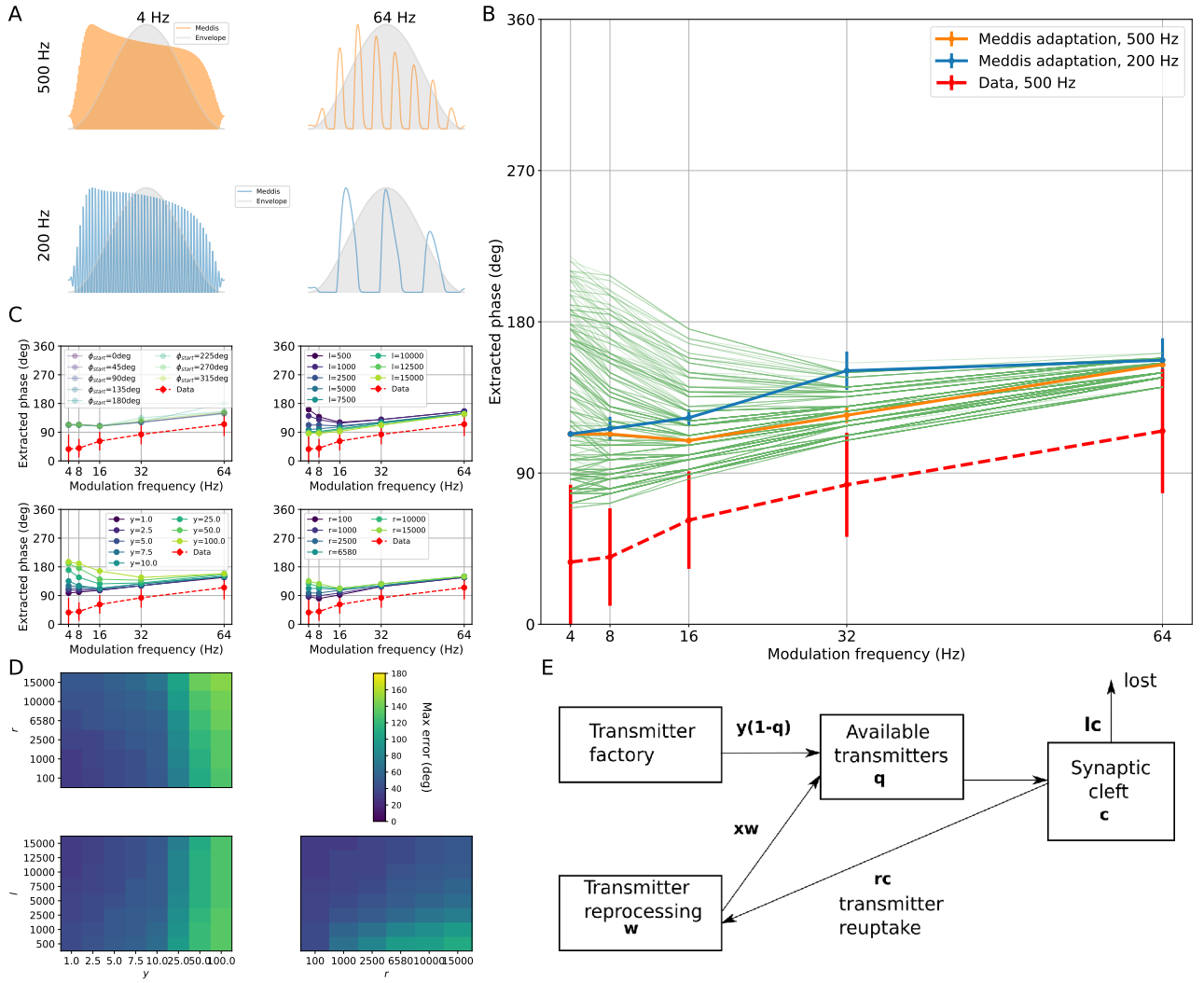

Figure 3.4: (A) Comparison between the envelope of the AMBB stimulus in (gray) and the sound after Meddis adaptation at 4 Hz and 64 Hz. In the first row sounds have a carrier frequency of 500 Hz. In the second row, sounds have a carrier frequency of 200 Hz. (B) IPD extracted by the model function of the modulation frequency with default parameter values (blue and orange) and additional phase plots with different parameter values at 500 Hz (green).(C) Behaviour of the model when changing the values of four main parameters: Starting phase ϕ_{start} of the stimulus (top left), the cleft loss rate coefficient l (top right), the replenishment rate coefficient y (bottom left) and the recovery rate coefficient r(E) Schematic of the Meddis model, adapted from Meddis (1988)

transmitters in the pool) is transferred from the reprocessing store to the free transmitter pool at each time step. Additionally, this reservoir reprocesses the quantity $rcdt$ of transmitters (c being the instantaneous quantity of transmitters in the synaptic cleft) that were not used at the synapse. The free transmitter reservoir releases a quantity $kqdt$ of transmitter in the synaptic cleft at each time-step. The coefficient k is a non-linear function of the instantaneous amplitude of the signal which models the non-linear behaviour of compression. An illustration

of the model can be seen in figure 3.4C

Methods The Meddis model was implemented in Python following the recommendations described in Meddis et al. (1990) which includes suggestions on the values of each parameters and implementation advice. Only the adaptation and non-linear compression part of the model were implemented. This model has been shown in Zhang and Carney (2005) to be equivalent to the model proposed by Westerman and Smith (1988). The signals were gammatone filtered (Hohmann, 2002) around the carrier frequency f_c and the output of the gammatone filter was subsequently used as input to the Meddis model (figure 3.3A). As in sections 3.1.1 and 3.1.2, the IPD returned by the model was chosen equal to the IPD maximizing the NCC coefficient between the left signal and the right signal (figure 3.3 and figure 3.4A and C). The implementation of the Meddis model used to run the simulations can be found in appendix 8.3. Overall the Meddis model follows the equations:

$$\frac{\mathrm{d}q}{\mathrm{d}t} = y(1-q) + xw - kq \tag{3.7}$$

$$\frac{\mathrm{d}c}{\mathrm{d}t} = kq - lc - rc \tag{3.8}$$

$$\frac{\mathrm{d}w}{\mathrm{d}t} = rc - xw \tag{3.9}$$

$$\tag{3.10}$$

The permeability factor $k(t)$ that dictates the instantaneous amount of neurotransmitters $q(t)dt$ being released in the synapse followed the equations:

$$\begin{cases} k(t) = \frac{g(I(t)+A)}{I(t)+A+B} \ \ if \ \ I(t)+A>0 \\ k(t) = 0 \ \ otherwise \end{cases} \tag{3.11}$$

Table 3.1: Meddis parameters

A	B	l	y	x	r
5	300	2580	5.05	66.31	6580

Results Using the default parameter values (table 3.1), passing the signals through the adaptation stage emphasized the onset of the sounds at all modulation and carrier frequencies (figure 3.4A). In addition, the strength of adaptation varied with modulation frequency, lower modulation frequencies led to a stronger adaptation (figure 3.4A). This emphasis resulted in IPDs extracted below 180° at all modulation and carrier frequencies (figure 3.2B) with the default parameters. The results at 500 Hz mostly reproduced the trend of increasing IPD extracted with modulation frequency (figure 3.4B), as seen in the data from Dietz et al. (2013). At 200 Hz, the IPDs extracted were generally above the ones at 500 Hz, which is qualitatively in line with the findings from (Hu et al., 2017). Varying the starting phase however did not lead to large variations in the results (3.4C top left) at 500 Hz, for this reason, the model was subsequently run with a single ϕ_{start} value equal to 0° in the rest of the results. Using the stimulus with a 500 Hz carrier frequency, we next varied individually the values of three of the main parameters (default parameter values can be seen table 3.1), the replenishment coefficient y, the recovery coefficient r and the cleft loss coefficient l (figure 3.4C). As expected, increasing l led to an increase of the strength of adaptation at all modulation frequencies as more transmitters are lost in the synapse. Conversely, increasing r or y led to more transmitters becoming available, which in turn resulted in higher firing rates and consequently in a weaker adaptation. Overall, no combinations of the parameters r, l and y tested led to fit of the data below 30°, the smallest error bar from Dietz et al. (2013) (figure 3.4D). However, the smallest absolute maximal error we obtained was close with a value equal to 32°. In general, Meddis adaptation still appears to be a good candidate to explain both the results from Dietz et al. (2013) and Hu et al. (2017) and we show in chapter 4 that a simplified version of the Meddis model can in fact account for the data.

3.1.3 Onset cells

Zilany model and Wang and Colburn

Higher up in the auditory system, cells that maximally respond to the onset of sounds have been observed (Rothman and Manis, 2003; Spencer et al., 2012, 2018; Wang and Colburn, 2012). Modelling onset behaviour was performed by combining a detailed model of the auditory periphery, the Zilany model (Bruce et al., 2018), with a model of LSO neurons (Wang and Colburn, 2012) specifically designed to replicate the onset behaviour exhibited by bushy cells (Rothman and Manis, 2003). To our knowledge, this combination of model has not been previously tested on the task described in Dietz et al. (2013).

Methods The onset model consisted of 6 different current channels. Namely, a leak current i_{leak}, a sodium current i_{Na}, a low-threshold potassium channel i_{KLT}, a high threshold potassium current i_{KHT}, an excitatory synaptic current i_{exc} and an hyperpolarization current i_h (equations 3.23). The hyperpolarization current implemented was identical to the one described in Rothman and Manis (2003). In addition, the inhibitory current originally present in Wang and Colburn (2012) was removed. The code of this onset model was implemented by Go Ashida and can be found appendix 8.3. Parameters relating to these current channels were set identical to the parameter values used in Dietz et al. (2016) and can be seen in more details in tables 3.2 and 3.3. A single channel of the stimulus was used as input to the Zilany model. The carrier frequency was chosen equal to either 500 Hz or 200 Hz. The output of the Zilany model was ,a peristimulus histogram (PSTH) binned at the sampling frequency (200 kHz), was used as input to the onset model which in turn returns a similar PSTH. Since a single trial at 500 Hz only produced a small amount of spikes (broadly 1 spike per modulation cycle), an accurate estimation of the extracted IPD using peak firing is not recommended. Instead, we repeated the simulations 100 times using 60 ANF channels. This process would be the equivalent of running the model for 100 onset cells and estimating the IPD from the population's response. We subsequently wrapped the PSTH around a single envelope cycle and returned the IPD maximizing the PSTH. The different parameters used in the model can be seen tables 3.2 and

3.3.

Table 3.2: Modified Wang model conductance and membrane parameters

C_s	g_{leak}	g_{KLT}	g_{KHT}	gNa
12 pF	20 nS	200 nS	150 nS	20 nS

Table 3.3: Modified Wang model reversal potential parameters

E_{leak}	E_K	E_{Na}	E_H
-65 mV	-70 mV	55 mV	-43 mV

All temperature parameters $T_{KLT}, T_{KHT}, T_{Na}, T_H$ were set equal to:

$$T_{body} = 37 \tag{3.12}$$

$$T_{op} = 22 \tag{3.13}$$

$$Q_{10} = 3 \tag{3.14}$$

$$T_{KLT}, T_{KHT}, T_{Na}, T_H = Q_{10}^{\frac{T_{body}-Top}{10}} \tag{3.15}$$

The current channels and the membrane potential followed the equations:

$$i_{leak} = g_{leak}(E_{leak} - v) \tag{3.16}$$

$$i_{KHT} = g_{KHT}(E_K - v)(w^4 z) \tag{3.17}$$

$$i_{KHL} = g_{KH}(E_K - v) \tag{3.18}$$

$$i_{Na} = g_{Na}(E_{Na} - v)(m^3 h) \tag{3.19}$$

$$i_H = g_H(E_H - v)r \tag{3.20}$$

$$i_{Exc} = 10^{-6} g_{Exc}(E_{ex} - v) \tag{3.21}$$

$$i_{tot} = i_{leak} + i_{Na} + i_{KHT} + i_{KLT} + i_{exc} + i_H \tag{3.22}$$

$$C_s \frac{dv}{dt} = i_{tot} \tag{3.23}$$

The gating variables w, z, n, p, m, h, r followed the equations below:

$$\tau_w(v)\frac{dw}{dt} = T_{KLT}(w_\infty(v) - w) \tag{3.24}$$
$$\tau_z(v)\frac{dz}{dt} = T_{KLT}(z_\infty(v) - z) \tag{3.25}$$
$$\tau_n(v)\frac{dn}{dt} = T_{KHT}(n_\infty(v) - n) \tag{3.26}$$
$$\tau_p(v)\frac{dp}{dt} = T_{KHT}(p_\infty(v) - p) \tag{3.27}$$
$$\tau_m(v)\frac{dm}{dt} = T_{Na}(m_\infty(v) - m) \tag{3.28}$$
$$\tau_r(v)\frac{dr}{dt} = T_H(r_\infty(v) - r) \tag{3.29}$$

The rate constants $w_\infty, z_\infty, n_\infty, p_\infty, m_\infty, r_\infty$ and their associated time constants $\tau_w, \tau z, \tau n, \tau p, \tau m, \tau r$ were implemented following:

$$w_\infty(v) = \frac{1}{(1+\exp(-\frac{v+48}{6}))^{\frac{1}{4}}} \tag{3.30}$$
$$\tau_w(v) = \frac{1.5+100}{6\exp(\frac{v+60}{6}) + 16\exp(-\frac{v+60}{45})} \tag{3.31}$$
$$z_\infty(v) = 0.5 + \frac{0.5}{1+\exp(\frac{v+71}{10})} \tag{3.32}$$
$$\tau_z(v) = 50 + \frac{1000}{\exp(\frac{v+60}{20}) + \exp(-\frac{v+60}{8})} \tag{3.33}$$
$$n_\infty(v) = \frac{1}{1+\exp(-\frac{v+15}{5})} \tag{3.34}$$
$$\tau_n(v) = 0.7 + \frac{100}{11\exp(\frac{v+60}{32}) + 21\exp(-\frac{v+60}{23})} \tag{3.35}$$
$$p_\infty(v) = \frac{1}{1+\exp(-\frac{v+23}{6})} \tag{3.36}$$
$$\tau_p(v) = 5 + \frac{100}{4\exp(\frac{v+60}{32}) + 5\exp(-\frac{v+60}{22})} \tag{3.37}$$
$$m_\infty(v) = \frac{1}{1+\exp(-\frac{v+38}{7})} \tag{3.38}$$
$$\tau_m(v) = 0.04 + \frac{10}{5\exp(\frac{v+60}{18}) + 36\exp(-\frac{v+60}{25})}) \tag{3.39}$$
$$h_\infty = \frac{1}{1+\exp(\frac{v+65}{6})} \tag{3.40}$$
$$\tau_h(v) = 0.6 + \frac{100}{7\exp(\frac{v+60}{11}) + 10\exp(-\frac{v+60}{25})} \tag{3.41}$$
$$r_\infty(v) = \frac{1}{1+\exp(\frac{v+76}{7})} \tag{3.42}$$
$$tau_r(v) = 25 + \frac{10^{-5}}{237\exp(\frac{v+60}{12}) + 17\exp(-\frac{v+60}{14})} \tag{3.43}$$

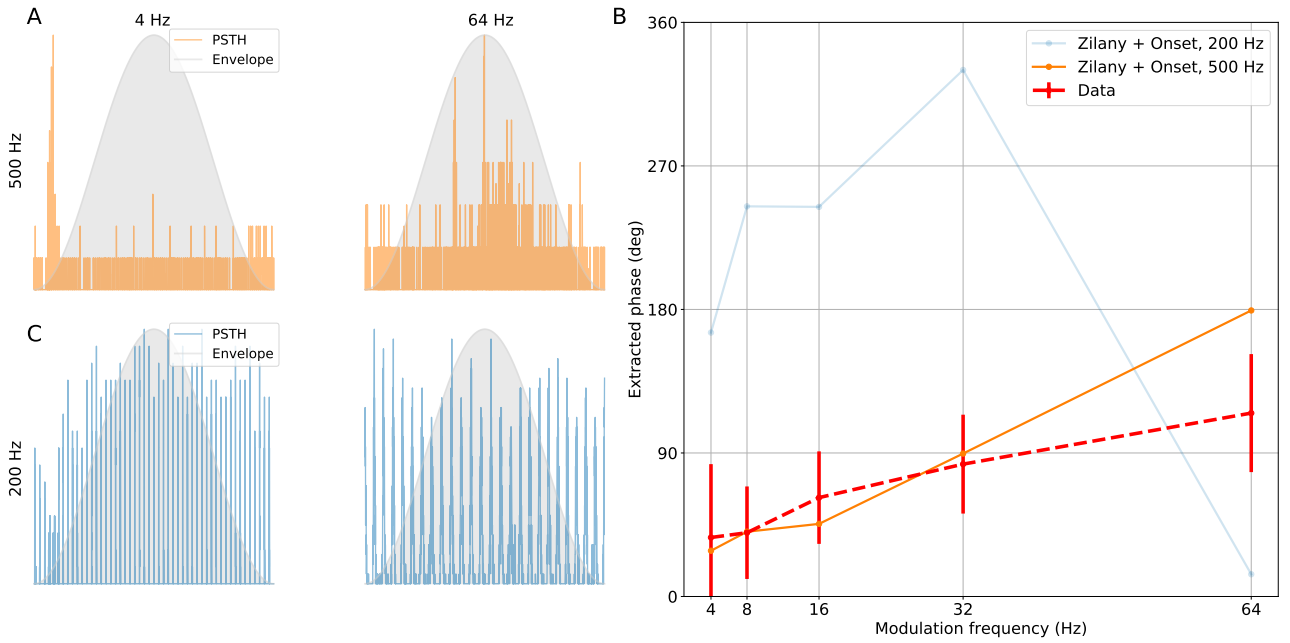

Figure 3.5: (A) Comparison between the envelope of the AMBB stimulus in (gray) and the peristimulus histogram (100 repetitions, 125 bins) at 4 Hz (left column) and 64 Hz (right column) at 500 Hz (top row) and 200 Hz (bottom row). (B) IPD extracted by the model as a function of the modulation frequency at 500 Hz and 200 Hz.

Results At 500 Hz, the model demonstrated sharp onset firing at low modulation frequencies (figure 3.5A top row) resulting in IPDs extracted early in the stimulus envelope (figure 3.5B). The $phase/f_m$ curve increased with modulation frequency in accordance with the data. However the fit of the data at 500 Hz was not satisfactory with the best fit producing an absolute maximum error equal to 64°. Conversely, at 200 Hz, the model fired much more regularly, at every cycle of the carrier frequency. This resulted in all IPDs being almost equally represented in the envelope (3.5B bottom row). This particular behaviour is interesting as it is in line with the findings that listeners do not prioritize the rising segment of the envelope when the carrier frequency is equal to 200 Hz (Hu et al., 2017). At 200 Hz however, the cell does not strongly favour a specific part of the envelope since the spikes are much more distributed. We investigate onset firing in more details in the next chapter (chapter 4).

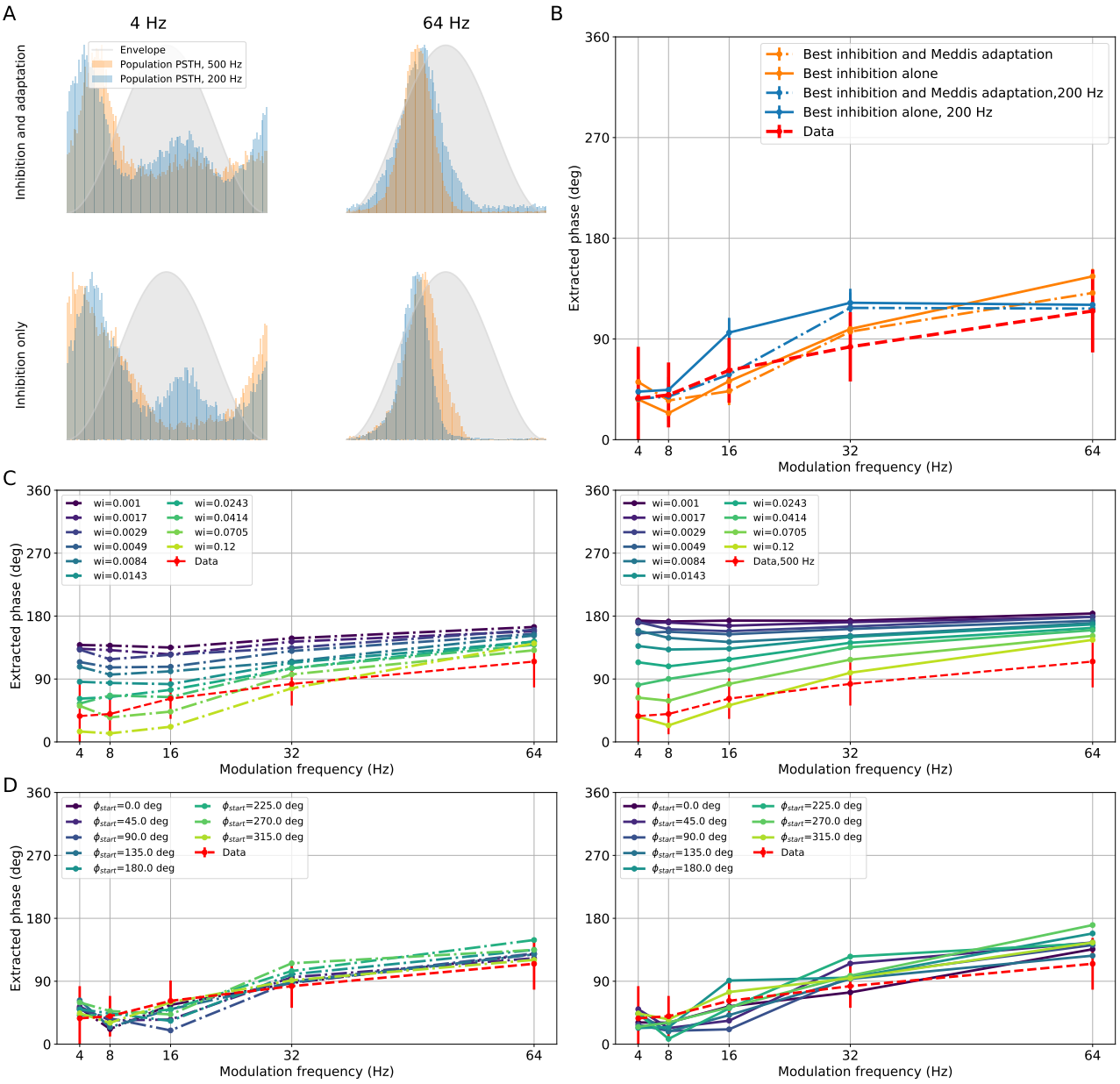

Figure 3.6: (A) PSTH of the network (20 repetitions, 125 neurons, $\phi_{start} = 0°$). Blue PSTH corresponds to a carrier frequency of 200 Hz. Orange distribution corresponds to a carrier frequency of 500 Hz. Bottom: Inhibition only. Top: Meddis and inhibition. (B) Best solutions with an without Meddis adaptation at 500 Hz and 200 Hz. Error bars are the circular standard deviation of the results over ϕ_{start} (C) Effect of the strength of the inhibition coefficient w_i on the extracted phase at 500 Hz. Each phase/f_m curve is the result of averaging circularly over all ϕ_{start} and all repetitions. Right: Inhibition only. Left: Meddis and inhibition. (D) Effect of ϕ_{start} on the extracted phase at the best values of w_i, at 500 Hz. The phase/f_m curves are the results of circularly averaging the IPD over all repetitions. Right: Inhibition only, $w_i = 0.12$. Left: Meddis and Inhibition, $w_i = 0.07$.

3.2 Central models

The previous models dealt with stages of sound processing that mainly occur before binaural interaction (where the inputs from the two ears are compared). These mechanisms principally affect the shape of the envelope. On the other hand, mechanisms performing lag-suppression located at higher stages, post-coincidence detection, are likely to be the results of neural interactions between neurons sensitive to interaural cues. Inhibitory circuits located in the medial superior olive (MSO) or in the inferior colliculus (IC) are good candidates for such computations. Here, we designed a network of IPD sensitive neurons capable of interacting with each other through long-range inhibitory connections. This type of connectivity results in the neurons in the network suppressing one another (unless they are close to each other in terms of feature distance). This arrangement is well suited to generate competitive dynamics between IPD tuned neurons.

Methods Two different networks were implemented. The first one included the same peripheral processing as in section 3.1.1. The second network included a Meddis adaptation stage similar to the one described in section 3.1.2 and with the default parameter values (table 3.1). In both cases, after peripheral processing, we modelled the spiking activity produced by the auditory nerve fibers (ANF) by using inhomogeneous Poisson processes whose rates were set proportional to the amplitude of the output of the peripheral processing. A refractory period equal to 0.5 ms was also added to the Poisson process. The spikes produced by the ANF fibers subsequently converged towards a network of coincidence detector (CD) neurons. The network was composed of 125 CD neurons. Because of the architecture of the model, each CD neuron responded best to one of the 125 delays (figure 3.3B). Coincidence detectors were modeled as Leaky Integrate and Fire (LIF) neurons. Inhibitory currents, originating from a subset of neurons in the network, were subsequently added to the input current of the neurons. Overall the network, was inspired by the network architecture described in Goodman and Brette (2010). The membrane voltage of the CD neurons followed the equations:

$$\begin{cases} \tau \frac{dV}{dt} = I - g_i - V + \sigma\sqrt{2\tau}\xi \\ \tau_i \frac{dg_i}{dt} = -g_i \end{cases} \tag{3.44}$$

In equations 3.44, V represents the membrane potential of the coincidence detectors, I represents the input provided by the auditory nerve fibers from both sides of the head and g_i is the inhibitory current which follows an exponential decay with time constant τ_i. The membrane time constant τ_m was chosen equal to 0.5 ms. This short time constant ensures that CD neurons only fire when spikes from both channels reach the neurons in a short time interval. The inhibition time constant τ_i was chosen equal to 10 ms. σ here refers to the standard deviation of the membrane potential without spikes while ξ is an additional gaussian noise satisfying the condition $< \xi(t), \xi(s) >= \delta(t - s)$. The inhomogeneous Poisson processes were further calibrated such that the CD neurons exhibit a mean (across neurons) firing rate of 50 Hz. For each trial (20 repetitions in each condition), the extracted phase was estimated by selecting the delay that maximized the firing rate. The resulting phases were subsequently circularly averaged over the 20 repetitions. For each neuron, neighboring neurons did not take part in the inhibition process, however, neurons located further away (in terms of IPD tuning) participated in it. This connectivity pattern can be summarized as:

$$\begin{cases} w_{i,j} = w_{j,i} = 0, \text{ if } \{|i-j| \bmod (N_{neurons}) < n \ \textit{and} \ |i-j| \neq 0\} \\ w_{i,j} = w_{j,i} = w_i \ \textit{otherwise} \end{cases} \tag{3.45}$$

In equation 3.45, w refers to the synaptic inhibitory weight and i and j represent the index of the coincidence detector neurons in the network and $N_{neurons}$ refers to the total number of neurons in the network. The value n corresponds to the number of neighboring neurons, around the neuron of interest, that do not participate in the inhibition process. The value of n was chosen equal to 14 here which represents a percentage of 77% of neurons in the network participating in the inhibition process. The inhibition connectivity was organized in a ring-shaped manner to address the circular nature of IPDs. A consequence of this ring-shaped arrangement is that

neurons tuned to IPDs close to 0° are considered direct neighbors to neurons with tuning close to 360°. The inhibition coefficient w_i was kept the same across all neurons.

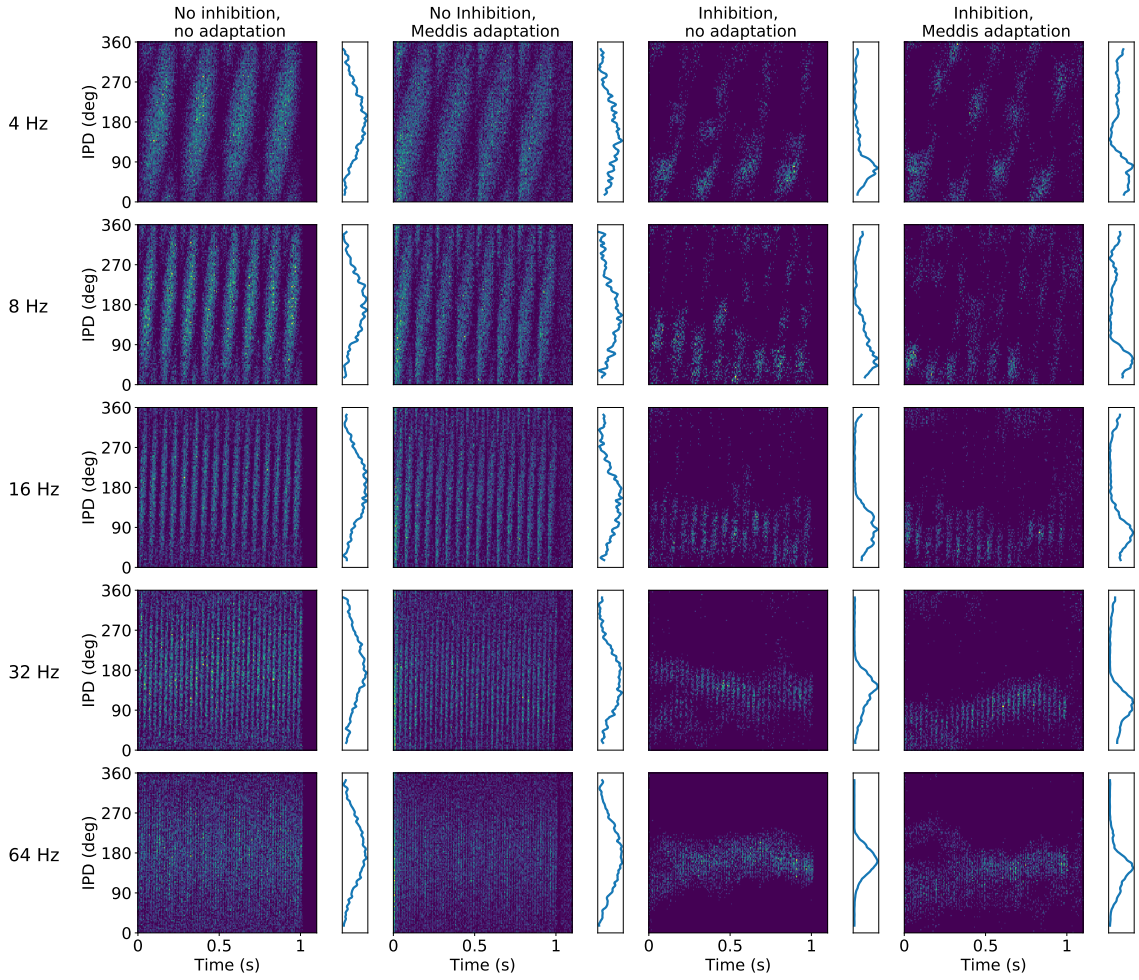

Figure 3.7: Density plots at 500 Hz. The densities are computed after 5 repetitions and the width of the bins was set equal to 5 ms. Rows: Modulation frequency. First column: Network without inhibition and without adaptation. Second column: Meddis inhibition and no adaptation ($w_i = 0.07$). Third column:Inhibition alone. Fourth column: Meddis inhibition and adaptation ($w_i = 0.07$). Right side of the density plots: normalized number of spikes per coincidence detector neurons. The curve was smoothed using a gaussian filter.

Results To investigate the ability of inhibition to account for onset extraction of temporal binaural cues, we implemented two different types of inhibitory networks. The first one includes a Meddis adaptation stage. This model is useful to study the combined effect of adaptation and inhibition on the results. The second network implements inhibition only. We ran the

model over 8 different ϕ_{start} ranging from 0° to 315°, 10 different values of the inhibition coefficient w_i and repeated the simulations 20 times, at 500 Hz, for each condition. The best fits obtained produced a maximum absolute error equal to 31° (obtained when $w_i = 0.12$) when only inhibition took place and 19° ($w_i = 0.07$) when Meddis and adaptation were both present (figure 3.6 B). We also ran the network at a 200 Hz carrier frequency (3.6B) using the inhibition coefficients that produced the best fits at 500 Hz ($w_i = 0.12$ without adaptation and $w_i = 0.07$ with adaptation). At both 500 Hz and 200 Hz, the phase/f_m curves mostly reproduced the pattern of increasing extracted phase with increasing modulation frequency (except at 4 Hz) as observed in the data from Dietz et al. (2013). In addition, at 500 Hz, and with or without Meddis adaptation, increasing inhibition led to a decrease of the phase extracted at all modulation frequencies (3.6C and D). The extent of the decrease highly depended on the modulation frequency. Lower modulation frequencies were more affected by inhibition (figure 3.6C and D). The reason for that is likely due to time-constant dynamics involved in the inhibition process. High modulation frequencies result in shorter envelopes which in turn leads to inhibition taking place at a later phase of the envelope. In both models, varying ϕ_{start} did not lead to large differences (figure 3.6D). This is likely due to the combined elements of the distribution of best delays (BIPD distribution) chosen to be uniform and of the connectivity pattern being arranged in a circular manner (equations 3.45). Finally, at 200 Hz, the phase/f_m curves were generally higher than at 500 Hz (figure 3.6B) which is in good agreement with the peak extraction at 200 Hz results uncovered in (Hu et al., 2017). Overall, the main effect of lateral inhibition was to create competitive dynamics between neurons tuned to different IPDs. The first neurons to spike start inhibiting cells further away, guaranteeing that their best IPD is most represented in the network. These spiking dynamics can be seen in the density plots in figure 3.7 where only a subset of neurons in the network respond to the stimulus when inhibition is present. In summary, long-range inhibition is a convenient mechanism to suppress lagging sounds and produced good fits of the data. An abstract version of this mechanism is investigated in more details in chapter 5.

3.3 Summary

We reviewed in this chapter a set of neural mechanisms that are good candidates to explain the early extraction of temporal binaural cues in amplitude modulated sounds. Some of these mechanisms take place at the periphery of the auditory system while others are located higher up in the auditory pathway. These mechanisms are adaptation, onset firing and inhibitory dynamics in a population of neurons. The best results obtained with adaptation led to phase/f_m curves below 180° and reproduced the pattern of increasing IPD with modulation frequency. The IPDs computed however, were generally extracted slightly later than the data at high modulation frequencies. This may suggest that other mechanisms are also taking place. Similarly, the onset model reproduced the trend of the data and exhibited an interesting property. At 500 Hz, spikes were mostly emitted at the onset of the envelope while at low frequency while, at 200 Hz, the model fired at the beginning of each carrier cycle leading to a much more distributed firing pattern. This results is in line with the finding from (Hu et al., 2017) that normal-hearing listeners mainly rely on the ITD located in the onset of the envelope for carrier frequencies above 200 Hz. Due to these different properties adaptation and onset combined together are likely to produce a good fit of the data (Dietz et al., 2013; Hu et al., 2017). This particular combination of mechanisms is examined more thoroughly in chapter 4, where we show that, in fact, each mechanism on his own can produce the data well. Other mechanisms described above are likely to take place in higher structures of the auditory system, medial superior olive or inferior colliculus. An example of such mechanisms is lateral inhibition circuits in networks of IPD sensitive neurons. Inhibition produced good fits of the data at 500 Hz and adding a Meddis adaptation stage to the model improved the fit even further (best error 19°). While being informative exploratory tools, the models described above present a certain number of limitations. First, we assumed in all models that the distribution of best interaural time delays was homogeneous within a certain frequency band. While this arrangement has been shown to exist in certain species (the barn owl for instance (Konishi, 2003)), it has not been observed in mammals so far (McAlpine et al., 2001). This issue is addressed in chapter 5. Second, because of the high number of parameters the models can exhibit, studying their behaviours in details

can be difficult. To address these issues, we designed in chapter 4 and 5, more abstract versions of each mechanism (adaptation, onset and inhibition) and study in greater details the range of behaviours each model can exhibit.

Chapter 4

Single cell behavior

The content of this chapter is partly adapted from Lestang and Goodman (2019).

Up until now, we have reviewed different mechanisms likely to be involved in the precedence effect and the early extraction of temporal cues in sounds. The models fell into three main categories: adaptation, onset and inhibition. In order to get a better understanding of each of these mechanisms we designed simple and general model of each of them and study their behavior in the next two chapter. Chapter 4 deals with adaptation and onset models while chapter 5 focuses on inhibitory dynamics.

4.1 Results

We investigate here 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 4.1), at a phase that increases monotonically with the modulation frequency but is always less than 180°.

In section 4.1.1, 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. 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. In this chapter, the input to the model was the envelope of the AMBB stimulus. The carrier frequency was otherwise added in 4.1.5. For more details see section 3.1.2.

4.1.1 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 4.1A). However, this peak occurs at 90° regardless of the 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.

4.1.2 Onset and adaptation

We model onset behaviour abstractly as the weighted difference of two low-pass filtered signals (with different time constants). The parameters are $\beta \geq 0$ the relative weight of the subtracted signal (so $\beta = 0$ corresponds to just low pass filtering the signal), and the time constants of

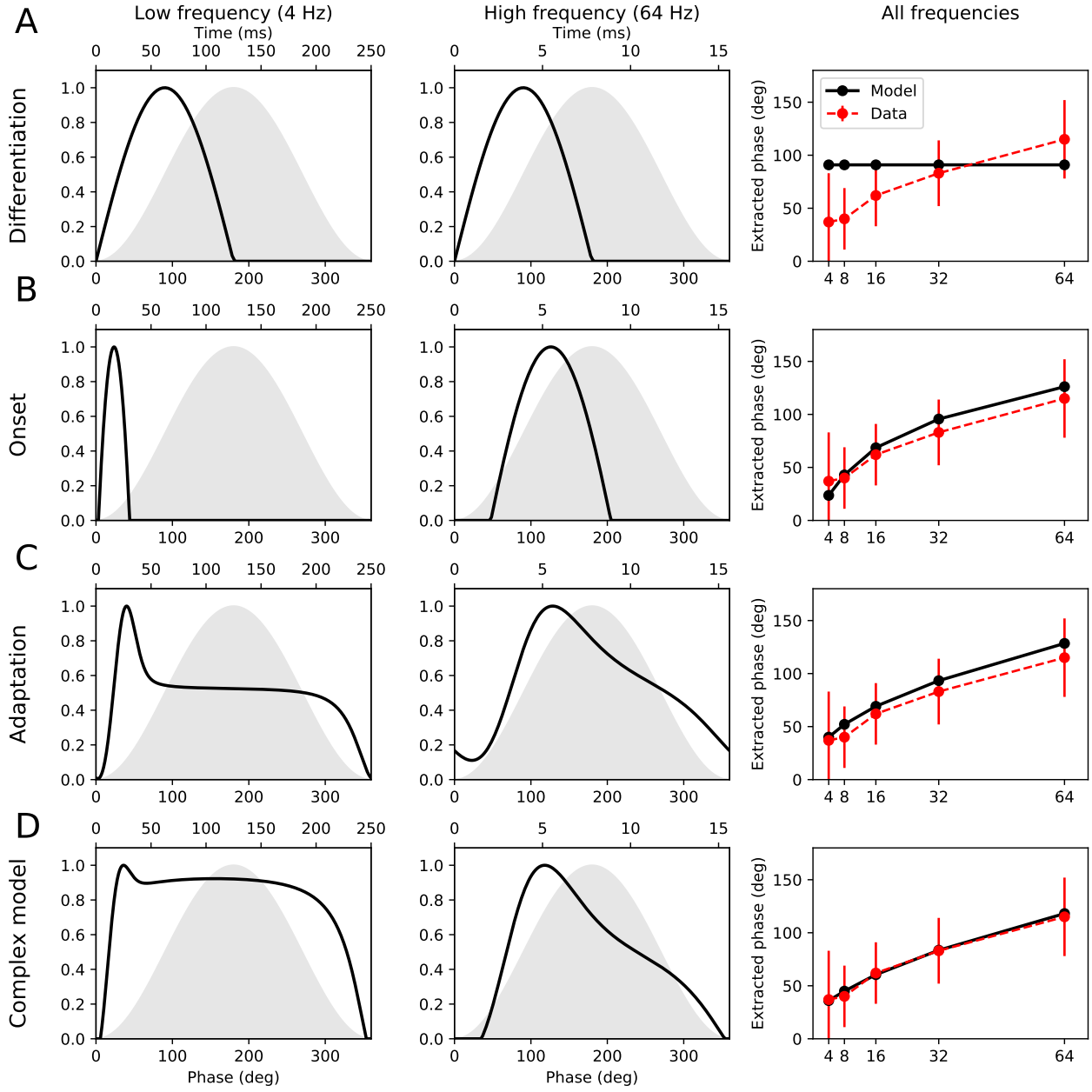

Figure 4.1: Basic mechanisms of the single neuron model. In each panel, the left and middle plots show, respectively, the response of the model at a low (4 Hz) and high (64 Hz) modulation frequency. The envelope is shown as the grey shaded area, while the output of the model is shown as a black line. 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 circular mean of 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.

the positive and negative filters (τ_e and τ_i respectively). 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. Our model is not spiking, but is a rate-based reformulation (see 4.2). With this model it is possible to closely fit the experimentally observed relationship between the modulation frequency and the extracted phase (figure 4.1B).

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 (similar to the Meddis model described in 3.1.2) that exhibits a response to a constant input that exponentially decays from an initial to a fully adapted response. 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). 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). Our adaptation model is therefore able to summarise a number of adaptation-related mechanisms, and uses only two parameters (see 4.2). Figure 4.1C shows that this adaptation mechanism is also able to closely fit the data.

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 4.1D). However, when we ran the model over a very wide range of different parameters (3-4M different parameter sets in each of the figures below), we saw that clearly a number of different sets of parameters work equally well. We explore this further in the next section.

4.1.3 Robustness of the mechanisms

The model has 6 free parameters and there are only 5 data points, so 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 4.2B), with the majority (73%) being monotonically increasing, for example. We investigate

this in more detail in 4.1.7 below. In addition, model fits are very robust, with large regions of the parameter space giving good fits (Figure 4.2A). 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 4.3). Using this model, we can determine which mechanisms are able to explain the data, and which mechanisms are essential.

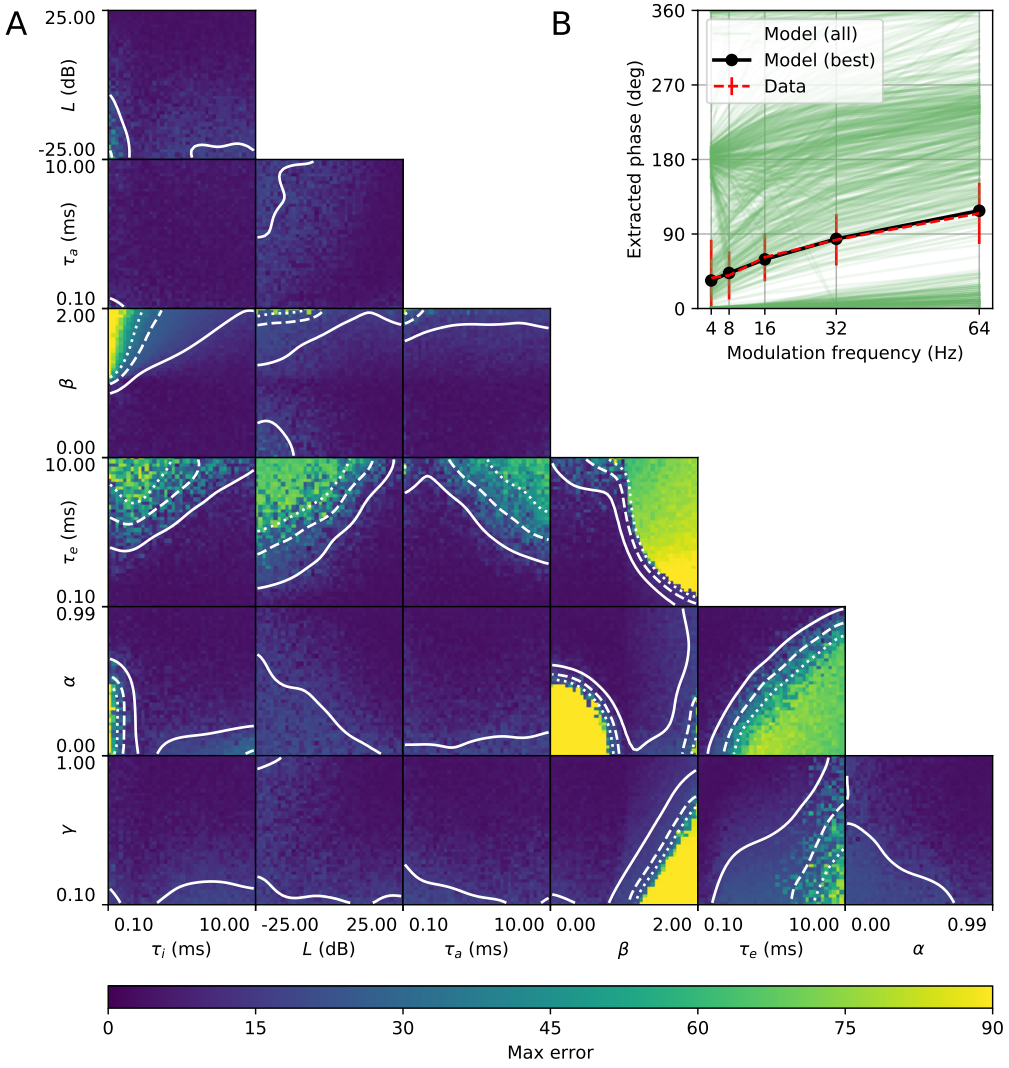

Figure 4.2: Parameter space. (A) Model error plotted as a function of different pairs of parameters. Each pixel shows the lowest error achievable for a particular fixed value for the corresponding pair of parameters. White contours show 15 degree error (solid), 30 degrees (dashed) and 45 degrees (dotted). (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).

We considered model variants where some of the parameters were given fixed values (Figure 4.2B) or some mechanisms removed (Figure 4.3). The key mechanisms of the model are adap-

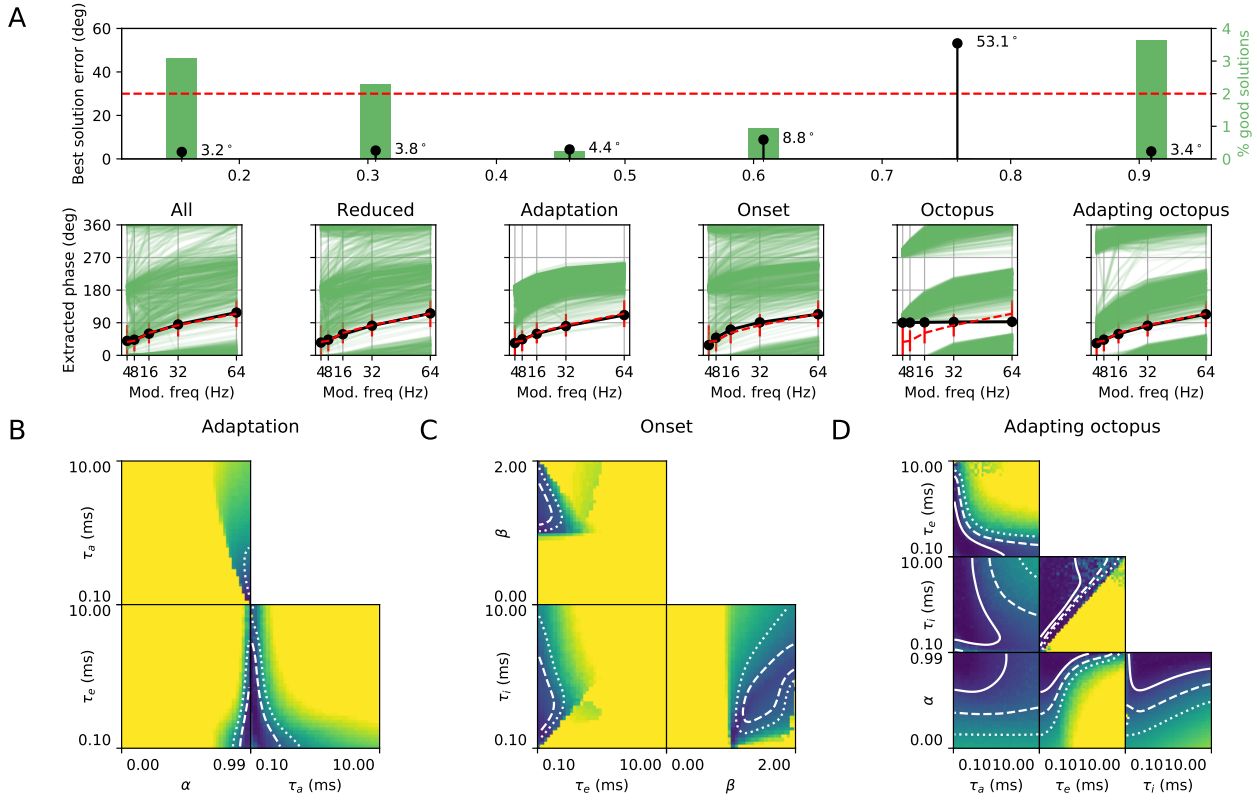

Figure 4.3: Comparison of restricted models. (A) Extracted phase curves (lower panels as in Figure 4.2B) for different model variants: *All* parameters available; a *Reduced* set where gain and compression were removed ($L = 0$, $\gamma = 1$), which is also the case in the remaining variants; *Adaptation* 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 4.2A for three of the model variants (adaptation, onset, adapting octopus).

tation 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 4.3ABC); and at least one of the two mechanisms must be present (Figure 4.2A, α versus β). The compression and gain mechanisms did not contribute substantially (Figure 4.3A). We also 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 4.3AD). For model variants including the onset mechanism, the positive or excitatory time constant τ_e must be small (Figures 4.2A and 4.4B), and the negative or inhibitory time constant τ_i must be larger than τ_e (Figure 4.3CD, 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 4.3A).

4.1.4 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 4.4), 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 (see section 4.2.3). 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 4.4A). 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 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 4.4B). Both cell types were about equally present in the parameter space considered.

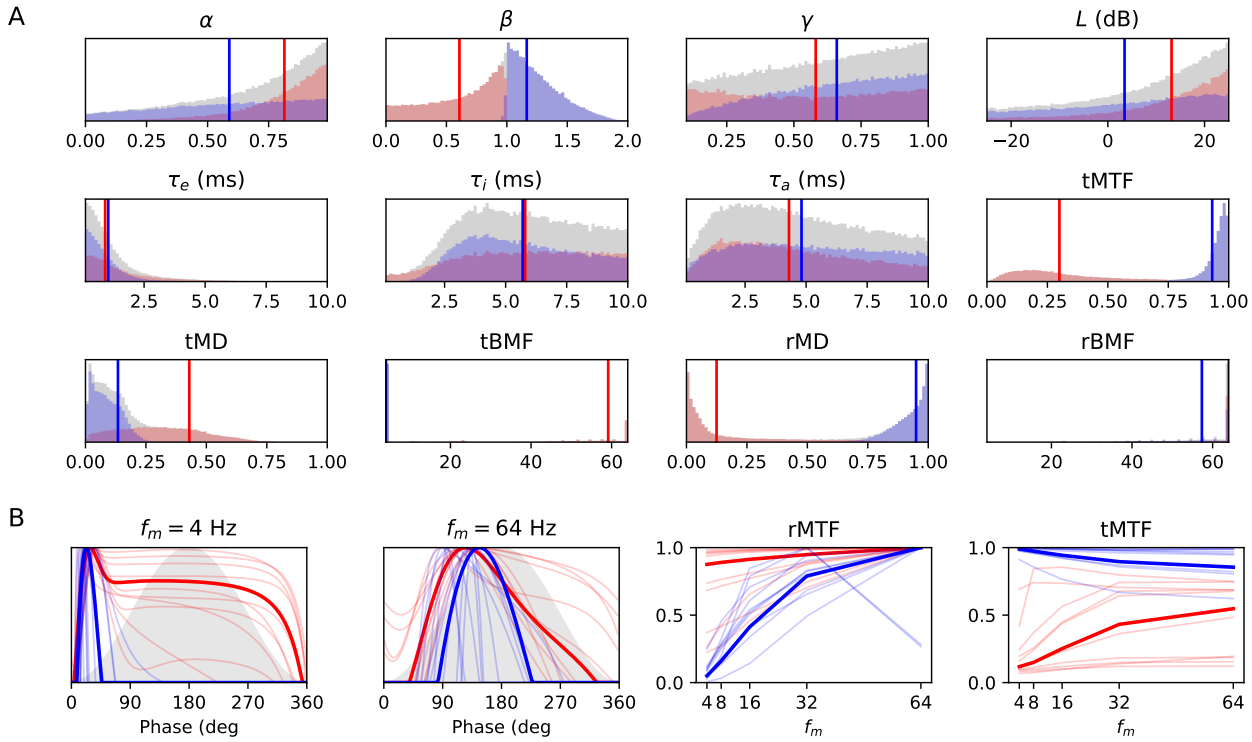

Figure 4.4: Cell types. (A) Histograms for all parameters and properties leading to an error of less than 30° (grey), for the strong onset cell class with tMTF$>$ 0.75 (blue), or the weak onset cell class with tMTF$<$ 0.75 (red). Vertical lines show mean values for the two classes. (B) Sample responses from the two cell classes. Left two panels as in Figure 4.1, 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).

4.1.5 Influence of carrier frequency

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), 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 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 4.2 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 4.5). 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 4.5B). 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).

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 (see section 6.3).

4.1.6 Effect of level

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

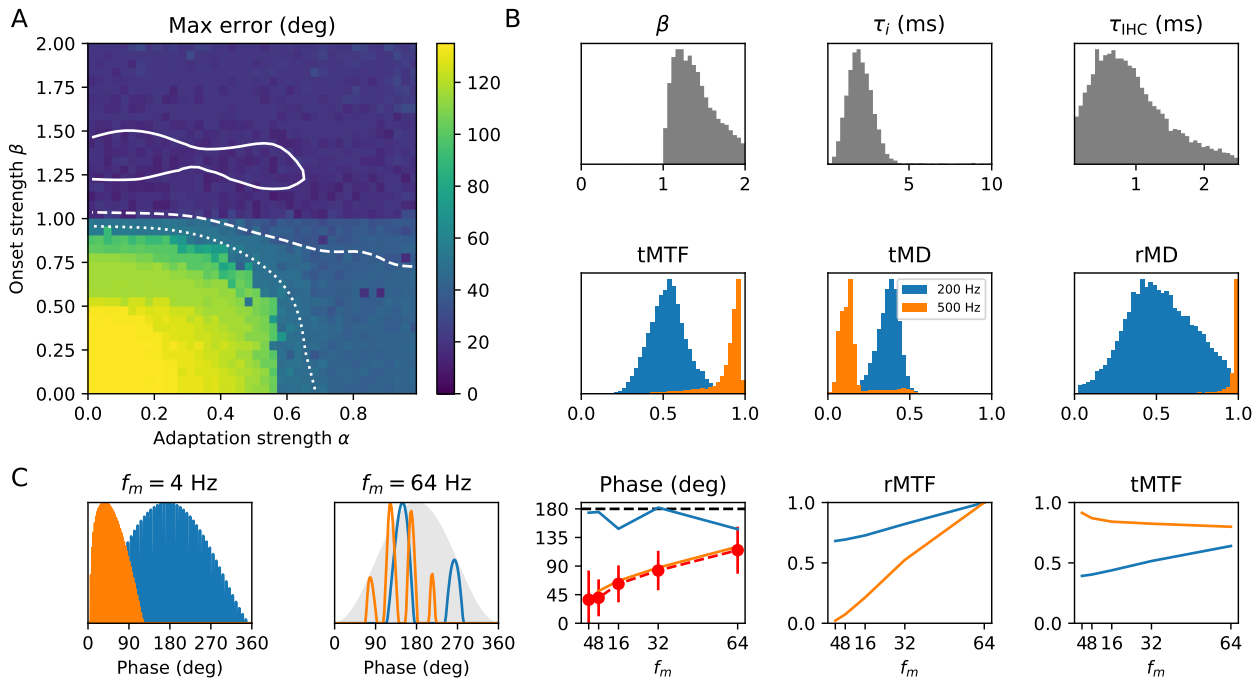

Figure 4.5: Effect of introducing carrier frequency. (A) Error in best fit for fixed adaptation strength α and inhibition strength β (as in Figure 4.4). Here, the error is measured at both $f_c = 500$ Hz and $f_c = 200$ Hz assuming a hypothetical flat 180° extracted in the latter case. The 200 Hz error is weighted less as there is no corresponding experimental data. (B) Parameter and property histograms as in Figure 4.4. Here, colours show properties at 200 Hz (blue) and 500 Hz (orange). (C) Best fit in detail as in figure 4.4.

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 4.6). 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).

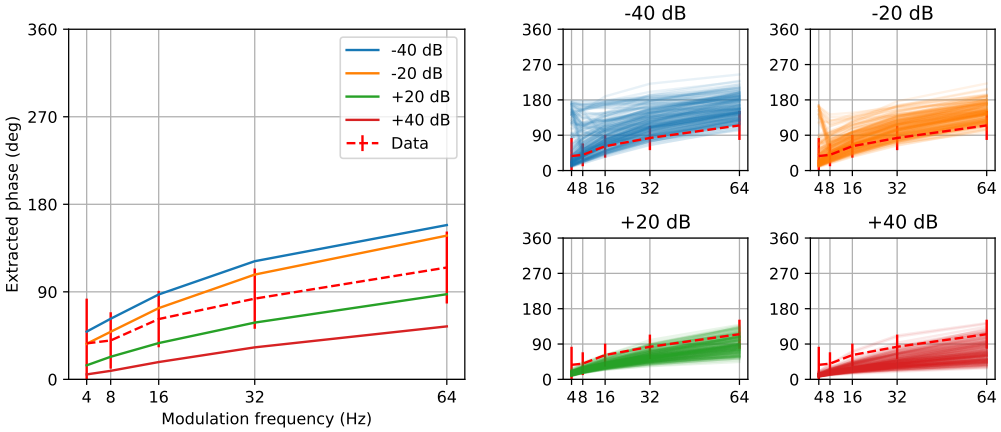

Figure 4.6: Single neuron model predictions for varying sound level. Left panel shows mean prediction at different sound levels for parameters that match the experimental data well at a sound level of +0 dB. Right panels show sample extracted phase curves for each sound level.

4.1.7 Curve fitting analysis

The models we used 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 from Dietz et al. (2013) if the data were different (figure 4.7). 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 IPD 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. The best fits obtained to the actual data were much better than the best fits for 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. Since the model was 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.

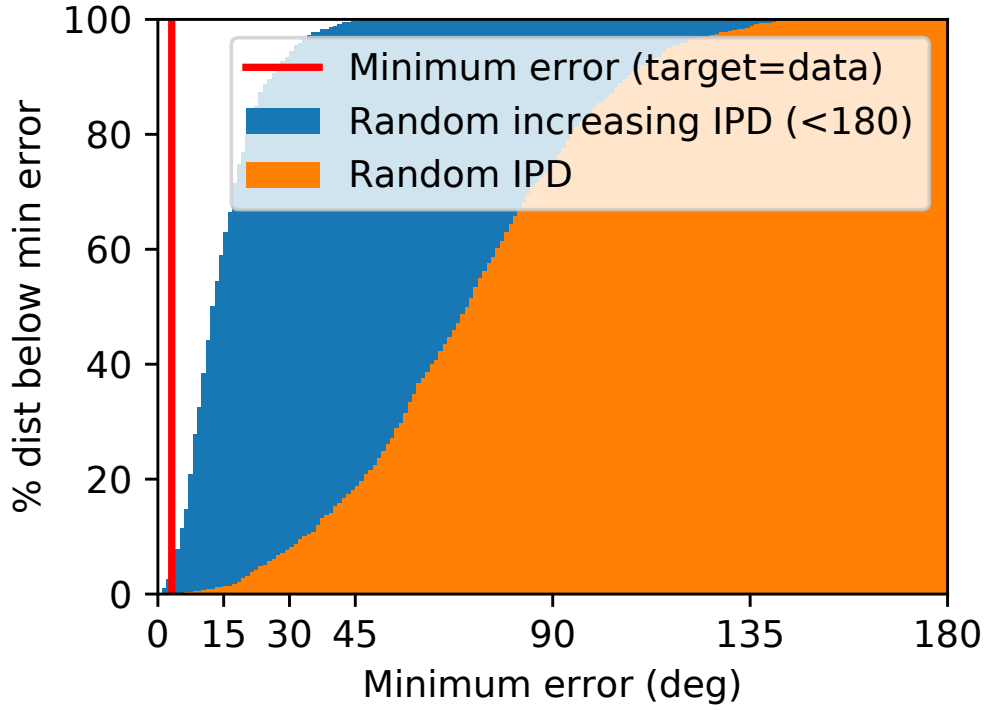

Figure 4.7: Single cell model. Best fits for experimental data (red line) and randomised target data (orange for fully random, blue when constrained to be monotonically increasing with modulation frequency less than 180°). Shaded areas show cumulative histograms of the minimum error found for the target 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°).

4.2 Methods

Models were implemented in Python using the Brian simulator (Goodman and Brette, 2008; Stimberg et al., 2014) 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.

4.2.1 Stimulus

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

$$E(t) = \tfrac{1}{2}(1 - \cos(2\pi f_m t)), \tag{4.1}$$

where f_m is the modulation frequency of the stimulus. In the first part of section 4.1 (up until section 4.1.5) 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 4.1.5) we add a carrier channel at 500 Hz.

$$S(t) = \cos(2\pi f_c t)\ E(t). \tag{4.2}$$

4.2.2 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

$$A_{\text{pre}}(t) = ([10^{L/20} S(t)]^+)^\gamma, \tag{4.3}$$

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). In the case where the carrier frequency was also added, the signal was passed through a first order low pass filter with time constant τ_{ihc}:

$$\tau_{\text{ihc}} \frac{\mathrm{d}A}{\mathrm{d}t} = A_{\text{pre}} - A. \tag{4.4}$$

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 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:

$$\frac{\mathrm{d}Q}{\mathrm{d}t} = -\kappa QA + \rho(1 - Q) \tag{4.5}$$

$$R_a = AQ. \tag{4.6}$$

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:

$$\tau_e \frac{\mathrm{d}R_e}{\mathrm{d}t} = R_a - R_e \tag{4.7}$$

$$\tau_i \frac{\mathrm{d}R_i}{\mathrm{d}t} = R_a - R_i \tag{4.8}$$

$$R = [R_e - \beta R_i]^+. \tag{4.9}$$

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:

$$\text{extracted phase} = 2\pi f_m \cdot \arg\max_t R(t)\ . \tag{4.10}$$

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:

$$\varepsilon = \max_{f_m} |IPD_{\text{data}}(f_m) - IPD_{\text{model}}(f_m)|. \tag{4.11}$$

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:

$$\varepsilon = \max\{\ \varepsilon_{500},\ \tfrac{1}{3}\varepsilon_{200}\ \}. \tag{4.12}$$

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).

4.2.3 Modulation measures

We computed several measures representing the variance of phase-locking and firing rates over modulation frequency.

Temporal modulation transfer function (tMTF) The mean tMTF was computed by averaging the value of the vector strength vs of the model's response over all modulation frequencies and shows the ability of the model to phase-lock to the envelope at all modulation frequencies:

$$\begin{aligned}
\phi(t) &= 2\pi f_m t && (4.13)\\
\overline{R(f_m)} &= \frac{1}{2\pi}\int_0^{2\pi} R(f_m, \phi(t))\, dt && (4.14)\\
vs(f_m) &= \frac{2\pi}{\overline{R(f_m)}}\sqrt{\left(\int_0^{2\pi} R(f_m, \phi(t))\sin(\phi(t))\, dt\right)^2 + \left(\int_0^{2\pi} R(f_m, \phi(t))\cos(\phi(t))\, dt\right)^2} && (4.15)\\
tMTF &= \frac{1}{N_{f_m}}\sum_{f_m} vs(f_m) && (4.16)\\
& && (4.17)
\end{aligned}$$

Equation 4.14 describes the process of averaging the firing rate across the duration of a modulation cycle. The mean firing rate is referred to as $\overline{R_(f_m)}$ in the rest of the section. Equation 4.15 shows the computation of the vector strength while equation 4.17 describes the computation of the mean temporal modulation transfer function.

Rate modulation transfer function (rMTF) The rate modulation transfer function is the result of averaging at each frequency the firing rate F of the model's response over time and dividing the result by the maximum mean firing rate over all frequencies. This measure

give an insight of the variation of the firing rate over the different modulation frequencies.

$$rMTF(f_m) = \frac{\overline{R(f_m)}}{\max_{f_m} \overline{R(f_m)}} \tag{4.18}$$

$$rMTF = \frac{1}{N_{f_m}} \sum_{f_m} rMTF(f_m) \tag{4.19}$$

In the set of equations above and in the rest of this method section, the overline symbol represents the operation of averaging temporally.

Temporal modulation depth (tMD) The tMD was computed by subtracting the value of the lowest vector strength coefficient to the highest one.

$$tMD = \max_{f_m} vs(f_m) - \min_{f_m} vs(f_m) \tag{4.20}$$

This measure quantifies the phase-locking distance between the best phase-locked response and the worst one.

Rate modulation depth (rMD) The rMD was computed by subtracting the lowest mean firing rate to the highest one and normalizing the difference.

$$rMD = 1 - \frac{\min_{f_m} \overline{R(f_m)}}{\max_{f_m} \overline{R_(f_m)}} \tag{4.21}$$

rMD measures the mean firing rate distance between the highest response and the lowest one.

Temporal best modulation frequency (tBMF) The tBMF was computed by extracting the value of the modulation frequency maximizing the vector strength:

$$tBMF = \arg\max_{f_m} vs(f_m) \tag{4.22}$$

Rate best modulation frequency (rBMF) The rBMF was computed by extracting the value of the modulation frequency maximizing the mean firing rate.

$$rBMF = \arg\max_{f_m} \overline{R(f_m)} \tag{4.23}$$

4.3 Summary

We found that adaptation and onset mechanisms led to an early extraction of the IPD information and produced excellent quantitative fits to the experimental data described in Dietz et al. (2013) across a wide range of parameters. We identified two main types of cells capable of accounting for the data: strong onset or strong adaptation. Onset cells displayed a preference for peak extraction with 200 Hz carrier frequencies, in accordance with the results published by (Hu et al., 2017), Our model also 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).

Chapter 5

Population model

The content of this chapter is partly adapted from Lestang and Goodman (2019).

5.1 Results

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 5.1).

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).

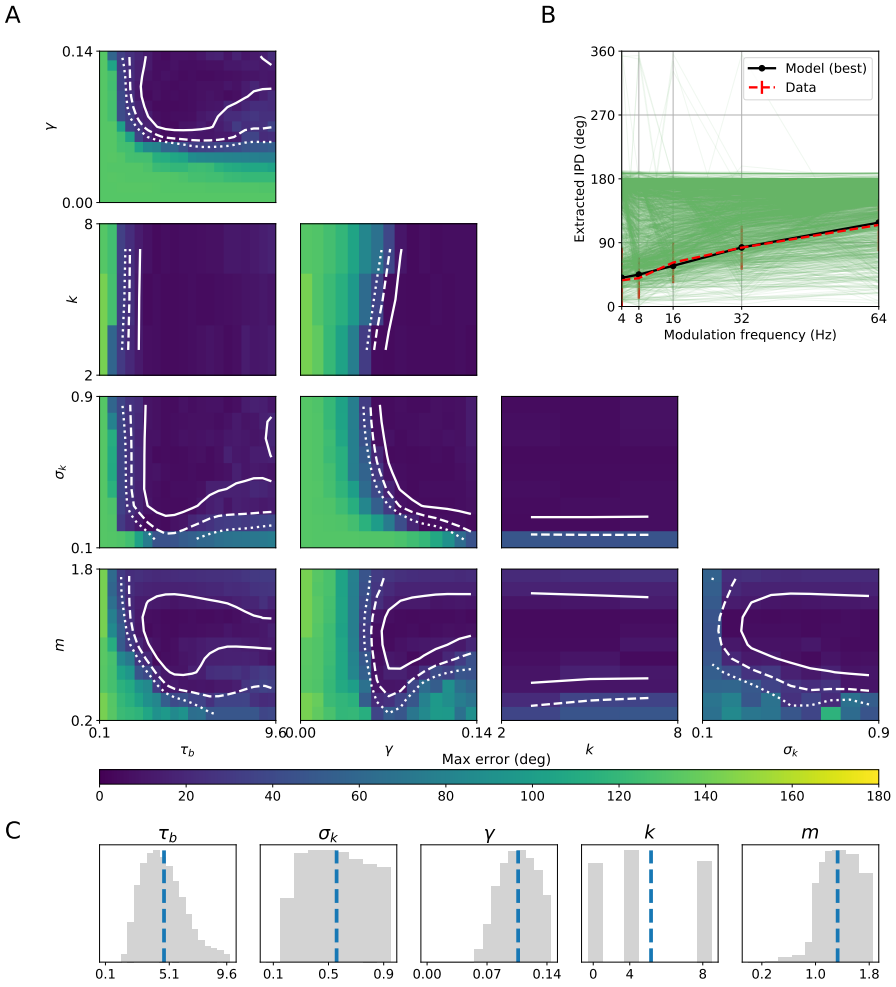

Figure 5.1: Population model parameter space, as in figure 4.2. (A) Model error plotted as a function of different pairs of parameters. Each pixel shows the lowest error achievable for a particular fixed value for the corresponding pair of parameters. White contours show 15° error (solid), 30° (dashed) and 45° (dotted). (B) Sample and best model results. (C) Histograms of the parameters leading to an error less than 30°.

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 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 4.2 for further details of the model.

5.1.1 Lateral inhibition and pattern matching

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 5.1B), and large regions of the parameter space led to low errors (figure 5.1A) 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 5.1C). There is a very steep drop in the number of good solutions where $m < 1$, suggesting 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 4.1.1 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.

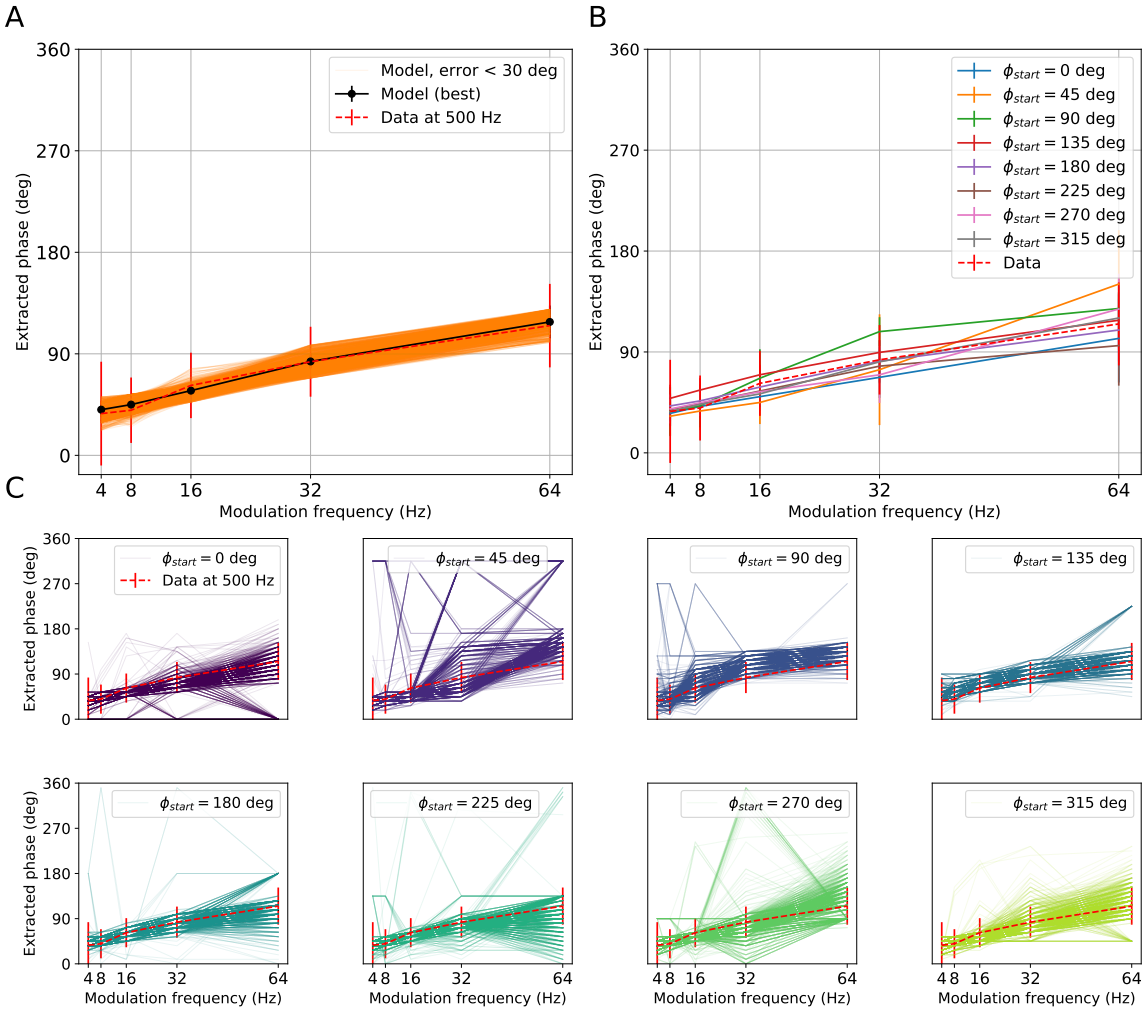

Figure 5.2: (A) Phase/f_m curves of all solutions within a 15 ° maximum error at 500 Hz. (B) Average over all best solutions with different ϕ_{start}. (C) Phase/f_m curves of all the best solutions with 8 different ϕ_{start}.Error bars show the circular mean across all ϕ_{start}

5.1.2 Effect of starting IPD

Since the BIPD distribution used in this chapter is not uniform, it is expected that the value of ϕ_{start} will produce different results. To study its effect on the results, we plotted for all best solutions within a 15° maximum error, the phase/f_m curves at each ϕ_{start} (figure 5.2). The shape of the best solutions mentioned above after averaging over all ϕ_{start} can be seen figure 5.2A. We found that ϕ_{start} did have an effect on the shape of the phase curves. For instance, at both 225° and 180° a subset of the phase curves tends to display an IPD extracted at 64 Hz that is much lower than at the other ϕ_{start} (figure 5.2). Similarly, at 45°, 90°, 225°, and 270°, a portion of the phase curve shows IPD extracted at much higher values (figure 5.2). However, because we use circular statistics, as in Dietz et al. (2013), these IPDs average over all phi_{start} to an IPD value within the data error bars (figure 5.2A and B).

5.1.3 Effect of the distribution of tuning

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 5.3). As before (section 4.1.5), 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 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 5.3D). For

carrier frequencies above 500 Hz, the preference did not shift any earlier.

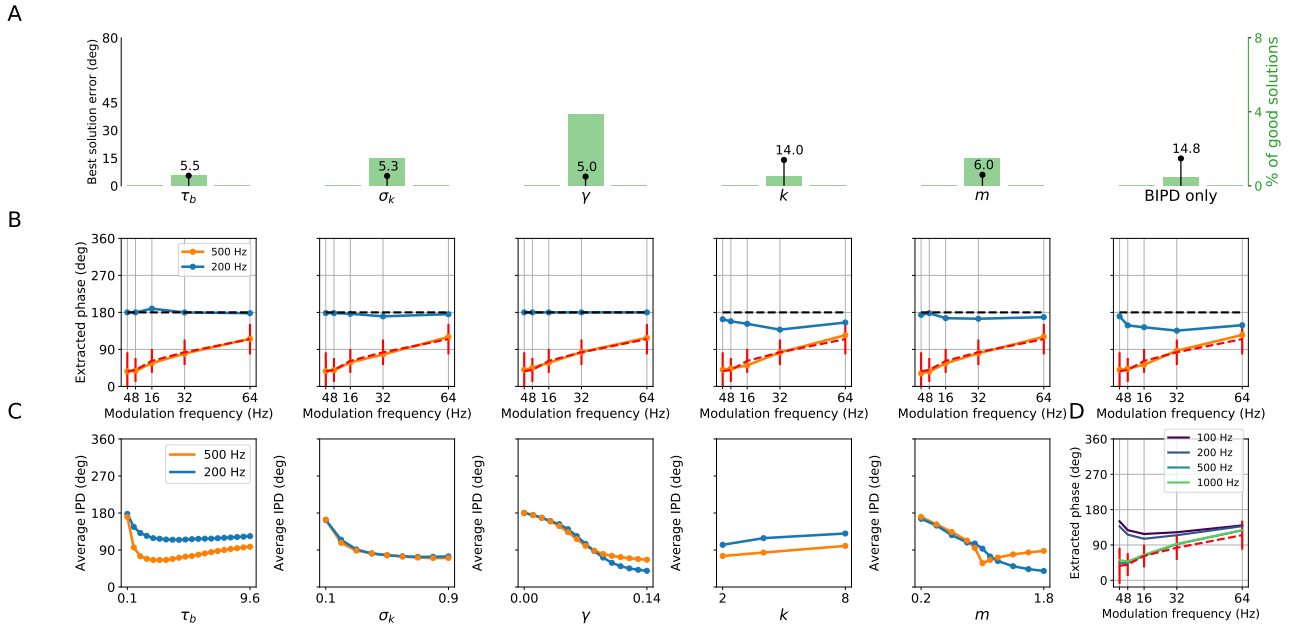

Figure 5.3: (A) Error in best solution (black lines and dots), and percent of the parameter space leading to good solutions (green bars), when allowing one parameter to take different values at 200 Hz and 500 Hz (except for "BIPD only" where all parameters are identical except for the BIPD distribution). (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 (see section 6.3). 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 5.3). 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 5.3C). 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.

5.1.4 Curve fitting analysis

Similarly to section 4.1.7, we examined whether or not our results obtained with the population model could be explained simply as curve fitting by looking at how well the model could fit the data if the data were different at 500 Hz (figure 5.4).The two same random drawing conditions described in section 4.1.7 were used. The best fits obtained to the actual data were much better than the best fits for 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 population model could fit the random data better than the experimental data only 8% of the time. Because the population model could not fit the majority of the random data, even when constrained to have a qualitatively similar shape to the experimental data, these results cannot simply be explained as curve fitting.

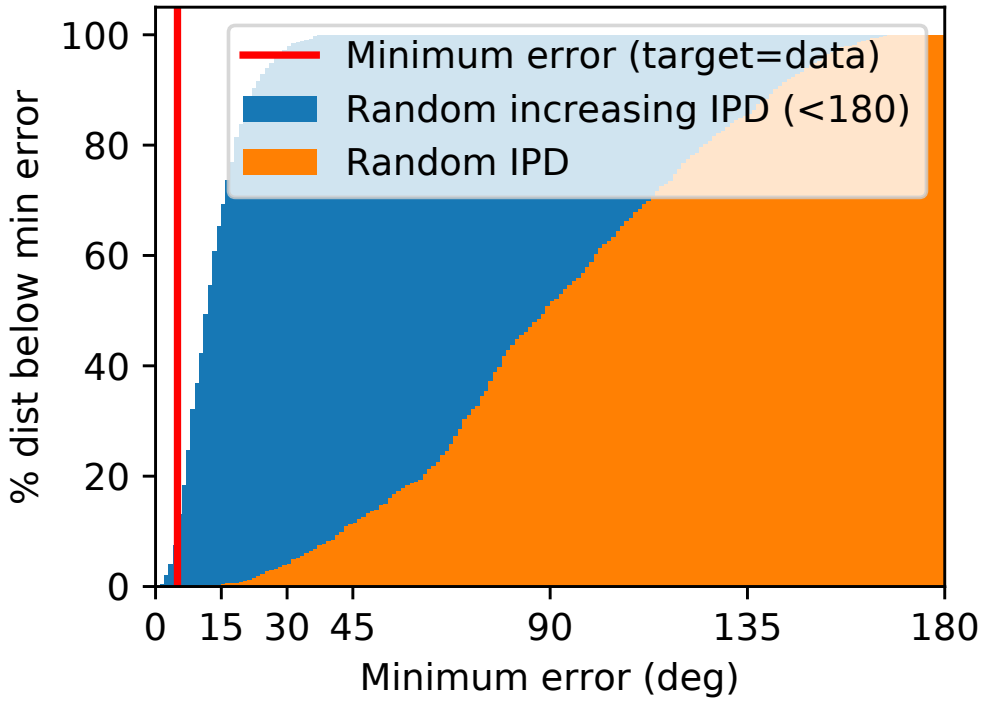

Figure 5.4: Population model. Best fits for experimental data (red line) and randomised target data (orange for fully random, blue when constrained to be monotonically increasing with modulation frequency less than 180°). Shaded areas show cumulative histograms of the minimum error found for the target 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°).

5.2 Methods

5.2.1 Modelling the neural response

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

$$T(\tilde{\phi}, \phi) = \left(\cos \frac{\tilde{\phi} - \phi}{2} \right)^k. \tag{5.1}$$

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 5.5C. 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

$$F(\tilde{\phi}, t) = T(\tilde{\phi}, \phi(t)) \cdot E(t). \tag{5.2}$$

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

$$E(t) = \tfrac{1}{2}[1 - m \cos \phi(t)]^+. \tag{5.3}$$

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:

$$T_{ij} = T(\tilde{\phi}_i, \phi(t_j)) \tag{5.4}$$

$$E_{ij} = \begin{cases} E(t_i) & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} \tag{5.5}$$

$$F = TE. \tag{5.6}$$

The pattern of activity generated by equation (5.6) can be seen in Figure 5.5A.

5.2.2 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:

$$d(\tilde{\phi}) = e^{-(\mu-\tilde{\phi})^2/2\sigma^2} + e^{-(-\mu-\tilde{\phi})^2/2\sigma^2}. \tag{5.7}$$

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°. BIPDs distributions at different frequencies can be seen in figure 5.5B.

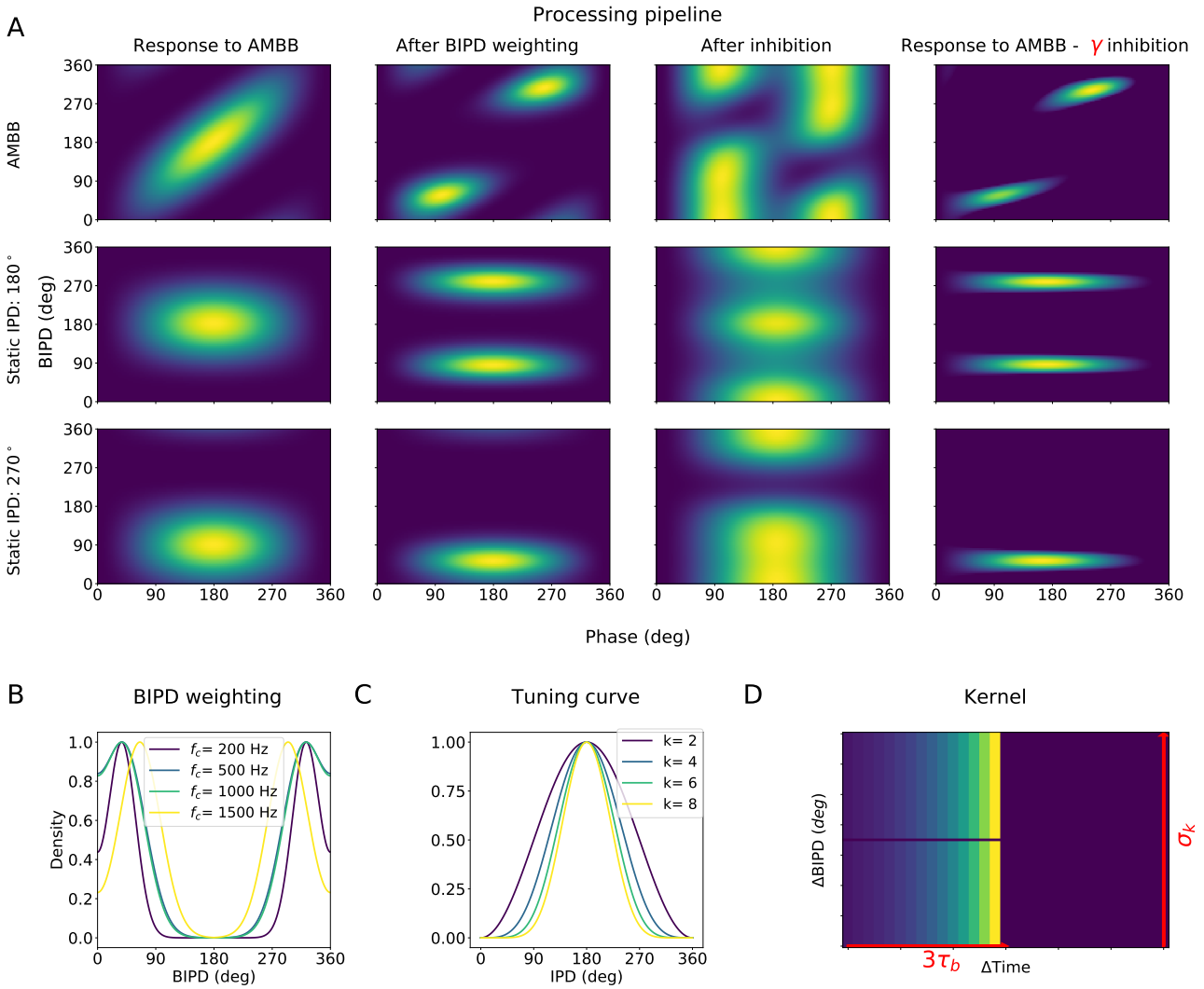

Figure 5.5: (A) Top row: the first column shows the response of the neurons to the stimulus. The response of the network after BIPD weighting is shown in the second column. The third column shows the inhibition layer, i.e. the network after inhibition. The last column shows the response of the network 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.

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 5.5A). 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:

$$D_{ij} = \begin{cases} d(\tilde{\phi}_i) & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} \tag{5.8}$$

$$F_W = DF = DTE. \tag{5.9}$$

5.2.3 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 5.6). 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:

$$i_{\text{inh}}(\tilde{\phi}, t) = \int_{-\infty}^{0} e^{\tau/\tau_b} F_W(\tilde{\phi}, \tau - t)\, \mathrm{d}\tau. \tag{5.10}$$

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

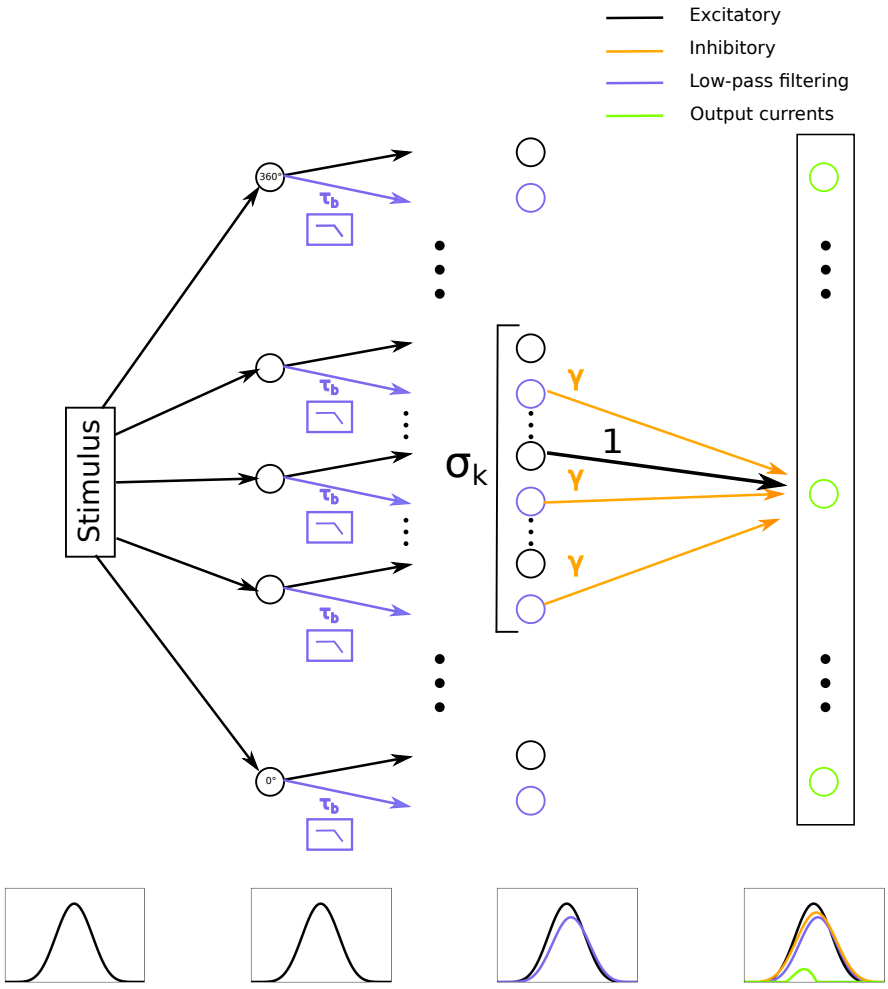

Figure 5.6: Structure of the network. Black circles represent the initial response of the neurons in the network to the AMBB stimulus. Each neuron is tuned to different BIPDs, from 0° to 360°. Blue circles represent the local inhibitory currents, and are low-pass filtered version of the black circles. 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.

inhibitory current is given by

$$I_{\text{inh}}(\tilde{\phi},t) = \sum_{\psi\in S_k(\tilde{\phi})} i_{inh}(\psi,t) \;=\; \sum_{\psi\in S_k(\tilde{\phi})} \int_{-\infty}^{0} e^{s/\tau_b} F_W(\psi,t-s)\mathrm{d}s, \tag{5.11}$$

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:

$$F_{WI}(\tilde{\phi},t) = [F_W(\tilde{\phi},t) - \gamma I_{\text{inh}}(\tilde{\phi},t)]^+. \tag{5.12}$$

From equations 5.10 and 5.11, 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 (5.12) can be rewritten as

$$F_{WI} = [F_W - \gamma F_W \circledast K]^+, \tag{5.13}$$

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

$$K_\phi(\psi) \;=\; \begin{cases} 1 & \text{if } |\psi| < 2\pi\sigma_k \\ 0 & \text{otherwise} \end{cases} \tag{5.14}$$

$$K_t(s) \;=\; \begin{cases} e^{s/\tau_b} & \text{if } s < 0 \\ 0 & \text{otherwise} \end{cases} \tag{5.15}$$

$$K(\psi,s) \;=\; K_t(s)\cdot K_\phi(\psi). \tag{5.16}$$

For the discretized formulation, we approximate K by cutting off $s < -3\tau_b$. With this approximation made, equation 5.13 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 5.5A.

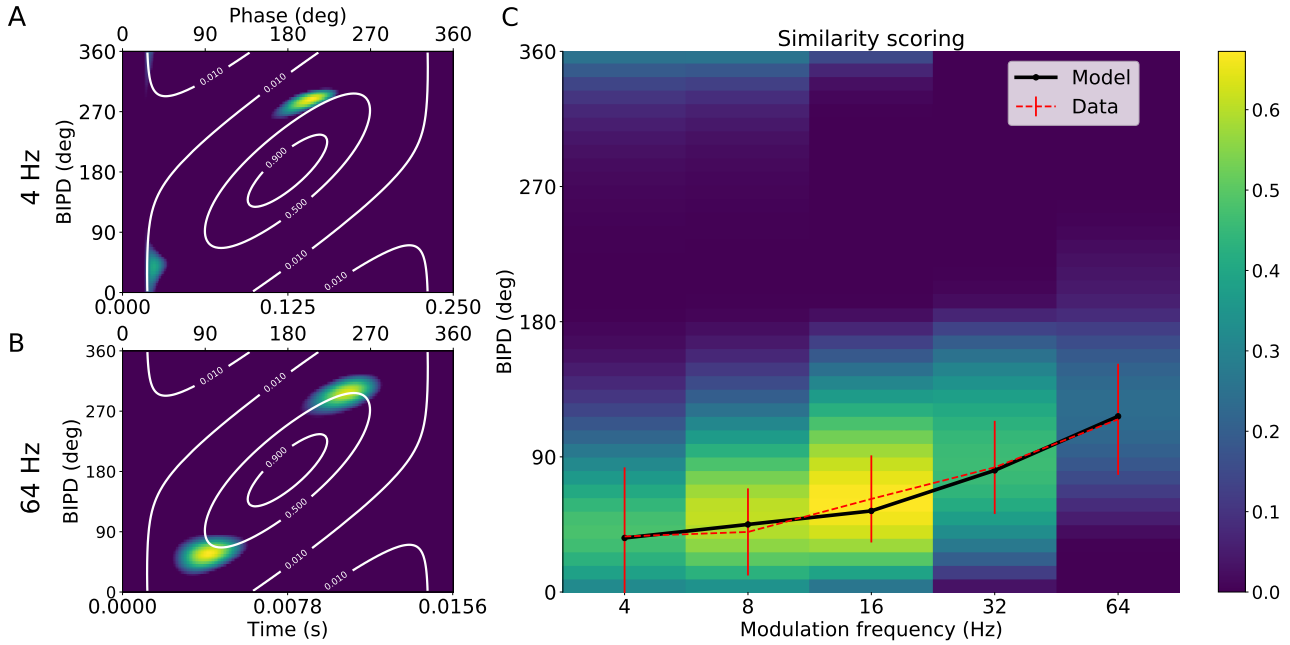

Figure 5.7: (A) Response of the network to the modelled amplitude modulated binaural beat (AMBB) stimulus with modulation frequency $f_m = 4$ Hz. Response of the network without inhibition is shown in white contours. (B) Same as A but with $f_m = 64$ Hz. (C) Similarity scoring. Similarity map showing 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.

5.2.4 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. The decoder we used is 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 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 5.5A 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 5.7C). This is equivalent to the cosine similarity metric as $|S_{WI}(\tilde{\phi})|$ is the same for all $\tilde{\phi}$.

$$\phi_{\text{out}} = \arg\max_{\tilde{\phi}} < S_{WI}(\tilde{\phi}), F_{WI} > \tag{5.17}$$

5.2.5 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}.

5.3 Summary

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. Using a lateral inhibition network, we found that we could obtain a great fit of the data described in both Dietz et al. (2013) and Hu et al. (2017). This model was able to reproduce good fits over a large range of parameters. Additionally, we found that the shift towards the peak of the envelope at low frequencies was produced by the difference in distribution of the interaural phase difference preferences of binaural neurons observed in mammals (McAlpine et al., 2001).

Chapter 6

The precedence effect and canonical mechanisms

The content of this chapter is partly adapted from Lestang and Goodman (2019).

6.1 Summary of findings

In chapter 3, we reviewed a set of neural mechanisms and studied their ability to account for the detection of interaural phase differences located in the slope of amplitude modulated sounds as reported in Dietz et al. (2013). Many of them demonstrated interesting behaviours, such as consistent extraction of the IPD below 180° and increase of the phase extracted with modulation frequency. These mechanisms were overall capable of emphasizing cues contained in the onsets of sounds, which has been suggested to be an important mechanism in responding to sounds following a direct path in a reverberant environment, and discarding later arriving information from indirect paths. This is likely an important factor underlying a number of phenomena, including spatial release from masking in cocktail party situations (Cherry, 1953a; Freyman et al., 1999). These mechanisms can be separated in three main groups: peripheral adaptation, peripheral onset firing and central inhibition. In an effort to design general models capable of reproducing the key behavior of each group of mechanisms, we designed and thoroughly tested

over a wide range of parameters abstract models of adaptation and onset firing (chapter 4) and lateral inhibition (chapter 5). In chapter 4, 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 reported 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 predicted that higher sound levels leads to a preference for earlier onsets (and vice versa), which is consistent with preliminary unpublished observations (Dietz, personal communication). Next, in chapter 5, we 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 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). It is less straightforward to predict the effect of level here because it could have multiple effects on the many cells involved. 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 experiment 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.

6.2 Generality of the mechanisms and canonical computations

The mechanisms modelled 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. Adaptation, for instance, 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; Yin, 1994), or higher up in the auditory pathway. In summary, in chapter 4 and 5 we modelled auditory neurons and populations with realistic distributions of IPD tuning, accounting for the frequency-dependent emphasis of rising versus peak sound energy in the perceived location of an ambiguous auditory stimulus. We found that a wide range of simple, general models based on neural mechanisms that are not specific to binaural processing – or even to the auditory system – all gave rise to the same effect. This suggests 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 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). Despite canonical circuits having been theorized to mainly exist in cortical microcolumns, our findings suggest that they could also be present in the midbrain. 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.

6.3 Neural response at 200 Hz and 500 Hz

In chapters 4 and 5, we questioned whether the same population of neurons respond to sounds at 200 Hz and 500 Hz. We investigate this topic below. At the level of the cochlea, incoming sounds cause the firing of auditory nerve fibers through the displacement of auditory hair cells. The activation of each auditory nerve largely depends on the spectral component of the sound and the preferred frequency of the auditory nerve. The sensitivity of the different auditory nerves to sounds with different frequencies has been shown to depend on the position of the nerve fiber in the basilar membrane (Von Békésy and Wever, 1960) and the level of the sound (Rhode, 1971). At a specific sound level, the degree of activation produced by one tone with a specific frequency on an auditory nerve with a specific preferred frequency has been studied in the past by using notched-noise maskers. The task consists in evaluating the signal detection threshold of a tone masked with broadband noise with different bandwidths around the frequency of the tone (Patterson, 1976; Weber, 1977; Moore and Glasberg, 1983, 1987; Moore et al., 1990) This procedure allows for the derivation of the shape of auditory filters (Patterson, 1976). Glasberg and Moore (2000) performed such an experiment while varying sound level. When the signal frequency was equal to 250 Hz, the auditory filters demonstrated a band-pass behaviour. Although the bandwidth of the auditory filters was shown to increase with level in the study, the difference in normalized gain at 250 Hz compared to 500 Hz was found larger than 20 dB even at high sound level (80 dB) (figure 7, Glasberg and Moore (2000)). The band-pass behaviour of the auditory filters was also confirmed in electrophysiological experiments. Recordings from the apex of chinchillas reported a band-pass

behaviour of the auditory filter with a best frequency equal to 500 Hz (figure 2, Cooper and Rhode (1997)). The study shows a clear decrease in sensitivity at 200 Hz compared to 500 Hz at multiple sound levels including 70 dB SPL and 80 dB SPL. In summary, at 75 db SPL, sounds at either 200 Hz and 500 Hz do cause the activation of the auditory nerve at each frequency but this activation is much greater at the frequency of the signal.

6.4 Future directions

The results described in chapters 4 and 5 mainly dealt with uncovering neural mechanisms capable of quantitatively reproducing the data described in Dietz et al. (2013). To understand the extent to which these mechanisms are taking place in the auditory system, we propose here some additional experiments.

6.4.1 Psychophysical experiments

The experiment by Dietz et al. (2013) can be easily modified to gain some insights on the mechanisms involved in producing the results of interest. For instance, changing the shape of the envelope would have a direct effect on the dynamics of inner hair cell adaptation. Indeed, fast adaptation leads to an enhanced sensitivity to variations in the amplitude of the stimulus. For instance, it would be possible to vary the value of the modulation depth of the stimulus to change the shape of the envelope. Joris and Yin (1992) showed that, in cats, increasing the modulation depth of the stimulus leads to an increased modulation of the auditory nerve response. Low modulation depth would therefore only produce small variations of the stimulus and would lead to a weaker adaptation (Joris and Yin, 1992). Consequently, we would expect the IPD to be extracted later with an AM stimulus with low modulation depth. For reference, similar experiments have been conducted in Klein-Hennig et al. (2011) and Dietz et al. (2016) with high frequency carriers. By presenting the AMBB stimulus with altered envelope shapes to human listeners we would be able to understand in more details the role of adaptation in the extraction of dynamic IPDs and compared it to modelled data. Second, our modelling

only covered a subset of the behaviours encompassed in the precedence effect. This subset is concerned with sounds lateralized with interaural time differences. The AMBB stimulus used in (Dietz et al., 2013) carries a time-varying interaural time difference while keeping an interaural level difference null. Coming up with a stimulus presenting opposite characteristics, a time-varying ILD and a null ITD, would be relevant to understand how our findings might transpose to sounds lateralized with ILDs. Such a stimulus can be designed by using two amplitude modulated tones with equal carrier frequencies and a phase delay equal to π between the two channels.

6.4.2 Electrophysiology and behavioral experiments

We uncovered in the previous chapters three main mechanisms capable of accounting for the data from Dietz et al. (2013), namely onset firing, adaptation and inhibition. A natural place to start to test the importance of these mechanisms experimentally would be to impair each of these mechanisms individually and observe the effect of these impairments on either behavioral results or neural activities. The mechanisms described in chapter 4, for example, could be tested in several ways. The stimulus could be played to anesthetized animals (cats, gerbils or ferrets) while recording the signal at the output of the posterior ventral cochlear nucleus (where octopus cells are located and direct input to the MSO and IC) or directly in the MSO or the IC. Changing the shape of the envelope for instance would lead to different adaptive dynamics and therefore produce different results compared to the control group (no alterations of the stimulus or impairments). Similarly, suppressing the activity of the octopus cells located in the posterior ventral cochlear nucleus could also be achieved by using optogenetics techniques. Optogenetics has been used in the past to stimulate the activity of the cochlear nucleus (Darrow et al., 2015) and was shown to be an efficient way to restore auditory behaviour in deaf gerbils (Wrobel et al., 2018). Our results at 200 Hz (figure 4.5) show that onset dynamics are necessary to obtain an estimated IPD around 180 degrees. Suppressing onset firing therefore should not allow the replication of the results by Hu et al. (2017). In summary, the experiment described above would be relevant in two aspects. First, by recording the activity of the cells in response to the

AMBB stimulus we would be able to identify, based on their modulation transfer functions, the presence or not of the two phenomenological cell types (onset type or adaptive type) described in chapter 4. Second, by impairing the mechanisms individually, we would also be able to quantify the importance of each mechanism in reproducing the results by Dietz et al. (2013) and compare it to our modelled data.

Next, to test the results produced by our population model (chapter 5), it would be conceivable to train mammals such as ferrets, guinea pigs or cats to perform the experiment described in Dietz et al. (2013) or if needed a simpler version of the experiment. Animals would be trained to accurately detect static IPDs in amplitude modulated sounds before being tested on the AMBB stimulus. In the case where it would in fact not be possible to train the animals on the exact task described by Dietz et al. (2013), a second approach could consist in recording the neuronal activities of a population of neurons in the structures of interest (MSO or IC), first in response to static IPDs and later in response to the AMBB stimulus. The extracted IPD would be estimated by simply running a pattern match decoder between the firing patterns obtained when the animals are listening to sounds carrying static IPDs and the ones produced by the AMBB stimulus. A first group, the control group, would be tested without any additional impairments to produce a baseline for the results, while over groups would exhibit a set of modifications designed to suppress the mechanism of interest. Since our population model relies heavily on inhibition, it would be relevant to disrupt inhibition by applying strychnine to the structure of interest. Strychnine has been used in the past to suppress inhibition in the MSO (Brand et al., 2002). Our model can predict that applying strychnine to the IC should lead to IPDs extracted later than the control group (figure 5.3C).

6.4.3 Modelling

Our simulations mostly dealt with amplitude modulated tones. However, it would be interesting to see how the models perform with a more diverse range of stimuli. For instance, it would be relevant to investigate how the models behave with stimuli of shorter duration and carrying a broader spectrum. Similarly to Xia et al. (2010), the models could be tested on pairs of clicks

and the results could be compared to the data from the psychophysical study related to the precedence effect by Litovsky and Shinn-Cunningham (2001). Additionally, our work does not directly address the performance of the models in response to more naturalistic stimuli but is likely relevant to understand speech perception in reverberant environments. This could be achieved by placing two microphones in the center of a reverberant room and playing speech segments from different locations. The recordings would later be used as input to the models. Alternatively, it would be possible to artificially lateralize the speech segments with ITDs and generate delayed versions of the speech segments lateralized to random locations.

Chapter 7

Auditory modelling framework

The content of this chapter is based on our study from Dietz et al. (2018). In this study, I played a major role in designing the framework and solely implemented it in Python. I also implemented other components of the framework including, pathway models and detector stages.

7.1 Motivation and challenges

Computational models are useful tools to interpret psychoacoustical results. However, comparing models and testing their performance on different experimental protocols can be challenging. We solved this problem in our approach in chapters 4 and 5 by having a single experiment and a general form of model with multiple mechanisms, enabling us to straightforwardly make comparisons. In the general case of model comparison across multiple experiments, this problem becomes more complicated. First, experimental data and computational models are not always available. Second, even shared models are not always version controlled which can be a limitation when trying to reproduce the results from a study (Eglen et al., 2017). In addition, models often vary in their implementation language which can complicate across-model comparisons. Third, models are often designed to meet the requirements of a specific experimental design. The type of input and output of the model is consequently largely decided by the nature of the experiment. This practice makes comparing models across multiple experimental designs

difficult. In order to facilitate these aspects of auditory modelling, we designed a framework (https://github.com/model-initiative/model_initiative) capable of integrating experimental protocols on the one hand and computational models, as artificial observers, on the other.

7.2 Architecture

Overall the framework is composed of four different components (figure 7.1B). The experimental program implements the experimental protocol which is in charge of the generation of the stimuli and of the interpretation of the response at each trial. On the side of the experiments, Waveform Audio File Format (WAV) files, storing the experimental stimulus are generated for each trial. The second part, the model interface, acts as an event listener for sound files generated on the experimental side. Once the stimuli are detected by the event-listener, the model interface launches the third component of the framework: the auditory pathway model. The model side in turn processes the sound and redirects its output to a decision stage. The newly produced decision is sent back to the side of the experiment which in response generates the stimuli for the new trial, in accordance with the updated decision. The role of the auditory pathway model is to process the incoming sound and as such to emulate the functioning of the auditory system during a psychoacoustical experiment. The pathway model takes as input the newly produced stimuli and produces a response which can take different forms (peristimulus histogram, cross-correlation coefficient, etc.). The model's responses to the different stimuli is later fed to the fourth component of the framework: the decision stage. The decision stage takes as input the different responses of the pathway model and returns a decision after each trial. Decision stages are overall the equivalent of human subjects making a decision based on their perception of the sounds and the nature of the task (which sound is louder, which sound appears to be located more on the left, etc.). Finally, the produced decision is written to a file and read by the experimental side to generate the stimuli for the new trial.

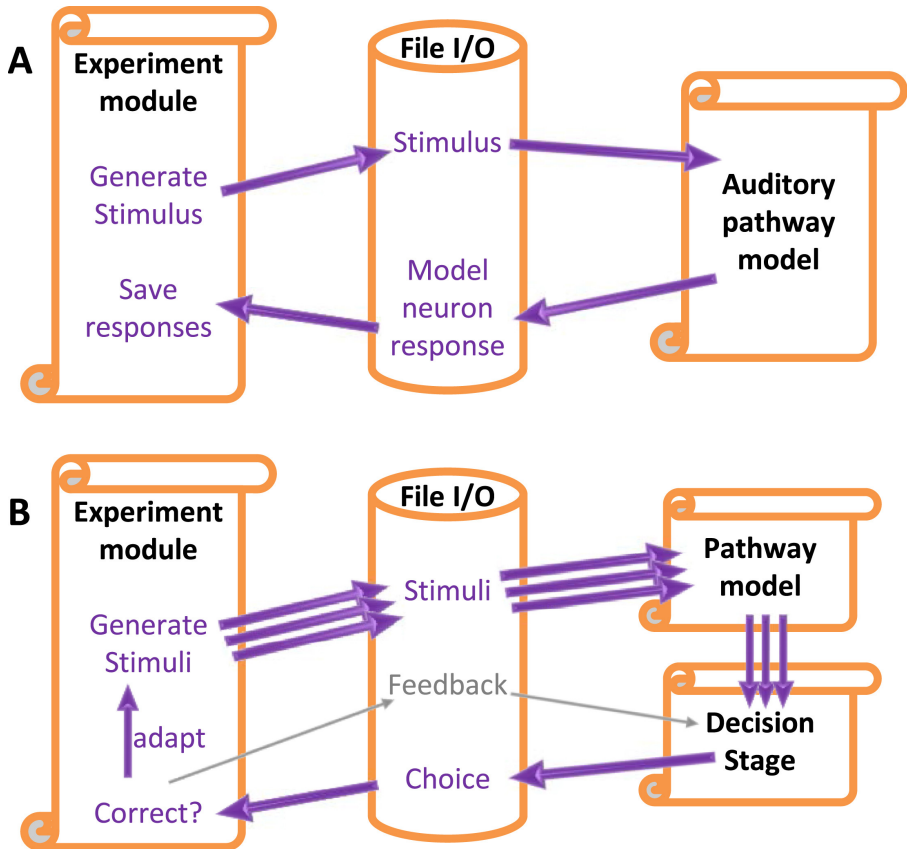

Figure 7.1: From Dietz et al. (2018),with permission as co-author. Functioning of the framework. (A) Functioning of the framework when studying neural response properties. (B) Functioning of the framework in an adaptive three-interval forced choice task. Each arrow represents a stimulus interval.

7.3 Benefits of the framework

7.3.1 The concept of the artificial observer

The concept of artificial observer (Dau et al., 1997a,b; Jepsen et al., 2008) is central to the functioning of the framework. Artificial observers are decision procedures that produce the same type of response to an experimental trial as a human subject would. This approach is useful as it allows models to be tested with the same experimental protocol used with human subjects (figure 7.2). In addition, experimental protocols often include adaptive staircase procedures. Adaptive procedures are experimental protocols in which the generation of the stimuli at each trial is largely dependent on the responses produced in the previous trials (Leek, 2001). The design of our framework was partly motivated by the need to take into account these adaptive procedures while testing computational models.

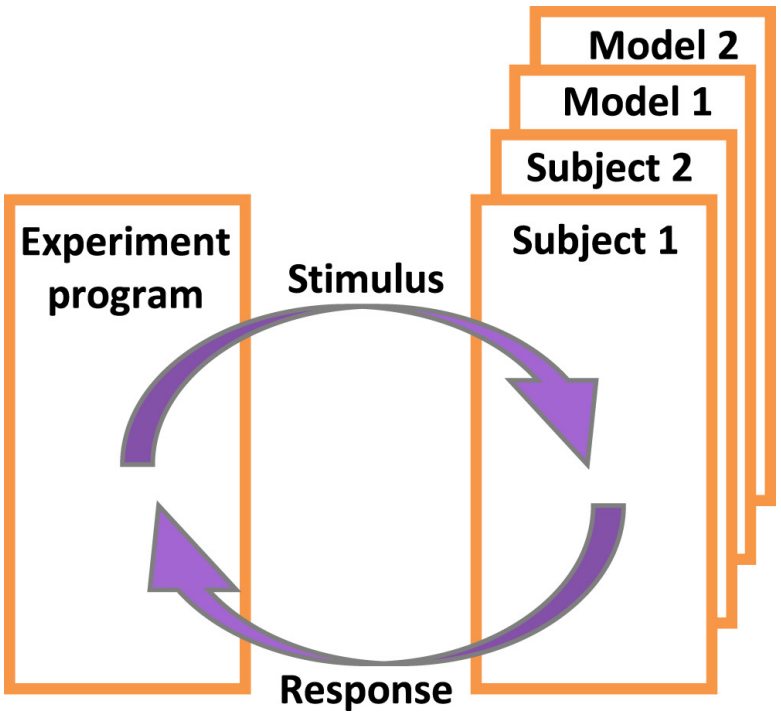

Figure 7.2: From Dietz et al. (2018), with permission as co-author. Artificial observer. Human subjects and models perform the experiment in the same manner ensuring direct comparison between human data and modelling data.

7.3.2 Separating pathway model from decision stage

Another important advantage of our approach is the separation of the pathway model from the decision stage. Decision stages are often task dependent and model dependent, which makes the comparison between models difficult. Separating pathway models from decision stages ensures that both stages of processing can be used in different tasks. This independence is established via the implementation of generalizable input/output conventions.

7.3.3 Language independence

The four different components of the framework are limited in their interaction to file exchange only. This property is useful as it allows models, detectors and experimental protocols to be written in different languages. Experiments, models and detectors are called via shell commands to safeguard the language independence of the framework. An important benefit of this approach is that it easily allows for the integration of third party software in all parts of the

framework. Examples of such third party libraries, included in the initial release of the auditory modelling framework, are given below:

- Brian hears, an auditory modelling library written in Python (Fontaine et al., 2011).
- A forced choice (AFC) toolbox (Ewert, 2013), a MATLAB/Octave modular framework to design adaptive psychoacoustical experiments.
- The Binaural cross-correlogram toolbox (Akeroyd, 2017) a MATLAB toolbox designed to facilitate the use of different cross-correlation functions over multiple frequency channels (Akeroyd, 2017).
- The Auditory modelling toolbox (AMT), a collection of MATLAB/Octave auditory models (Søndergaard and Majdak, 2013).

7.4 Experiments, pathway models and decision stages implemented

The framework also consists in a collection of experiments, pathway models and decision stages. Some examples are listed below.

7.4.1 Experiments

A set of experiments were added to the initial release of the framework. The protocol of these experiments ranged from just noticeable detection of ITDs tasks to signal in noise detection BMLD procedures and absolute localization tasks. All experiments were implemented in Matlab. Two experiments were implemented using the AFC toolbox (Ewert, 2013). The first one implements the experimental design described in (Klein-Hennig et al., 2011) which investigates the effect of the shape of the envelope in ITD detection in high frequency sounds. The second

one corresponds to the BMLD protocol detailed in van de Par and Kohlrausch (1999). Additionally, just noticeable ITD experiments from Bernstein and Trahiotis (2002) and Dietz et al. (2015) were also added to the framework.

7.4.2 Pathway models

For each experiments detailed in section 7.4.1, we implemented the corresponding modelling approach used in the study. We also added the model described in Breebaart et al. (2001) using the AMT Matlab implementation (Søndergaard and Majdak, 2013) and a Python implementation of a spiking model inspired by (Goodman and Brette, 2010).

7.4.3 Detectors

A centroid lateralization detector from the binaural cross-correlogram toolbox (Akeroyd, 2017) was also released with the package as well as the detector stage used in (Breebaart et al., 2001) from the AMT toolbox.

7.5 Summary

We introduced a framework whose main goal is to facilitate the reproducibility of auditory modelling studies by encouraging users to share their experiments and model. In addition, the structure of the model simplifies the process of comparing models on different experiments by separating pathway models and decision stages and by allowing experiments, models and detectors to be implemented in different languages.

Chapter 8

Conclusion

8.1 Summary of Thesis Achievements

Our work focused on understanding the computation of interaural time differences in low-frequency sounds when confounding temporal localization cues are present. We uncovered in chapter 3, a set of neural mechanisms capable of extracting interaural time differences in the rising portion of sounds. These mechanisms were detailed models of spike frequency adaptation, onset firing and inhibitory dynamics. In order to examine in more details the behaviour of each mechanisms, we designed abstract models of each mechanisms and studied their performance over a wide range of parameters. In chapters 4 and 5 we demonstrated that we were able to account for the data reported in Dietz et al. (2013) with general models of adaptation, onset firing and inhibition, commonly found in the brain. This result suggests that the precedence effect might be the result of exaptive mechanisms rather than specialized circuits (chapter 6). Overall, the models described were mainly tested on two different experiments (Dietz et al. (2013) and Hu et al. (2017)). Testing and comparing models on a wider range of experiments can be a significant challenge as the decision stage of the model fully depends on the nature of the experiment. As a mean to facilitate the comparison of computational models for auditory modellers and encourage reproducibility in the auditory field, we also described in chapter 7, a framework capable of integrating experimental protocols on one side and computational

models on the other. In this framework, models and decision stages are independent one another which facilitates the use of models (regardless of their implementation language) on different experimental paradigms and simplifies the integration of third-party libraries in the framework.

8.2 Applications and future work

First, our findings are likely to be relevant in the field of cochlear implant design. Cochlear implant users generally show poorer performance on sound localization tasks compared to normal hearing listeners (Grantham et al., 2008). This is partly due to a poorer sensitivity to temporal cues in CI users. Second, in a cocktail party scenario, the location of the speakers are important cues to segregate the different incoming sounds and focus on a particular conversation. Our work on robust interaural time differences computation in the presence of competing stimuli is likely to improve our understanding of auditory scene analysis.

8.2.1 Cochlear implants (CI)

Cochlear implants bypass most of the periphery of the auditory system to directly excite the auditory nerves. Consequently, inner hair cell adaptation, amongst other mechanisms, does not take place in CI users. A recent study has shown that integrating hair-cell adaptation in cochlear implants improved speech perception in acoustic environment with mild background noise (Azadpour and Smith, 2016). Adding an inner hair cell adaptation stage to cochlear implants is likely to also improve sound localization abilities in bilateral CI users and might help in establishing precedence effect related behaviours in CI users. Indeed, Brown et al. (2015) showed that it is possible to improve lead-lag fusion in CI users by attenuating the amplitude of the lagging sound the same way inner hair cell adaptation does. Testing whether inner hair cell adaptation could lead to improvements in sound localization in the presence of reverberations in cochlear implant users could consist in setting up an experiment similar to the one described in Kerber and Seeber (2012). The experiment would be performed in free-field conditions with multiple speakers surrounding the subject. The subject would aim at localizing a target sound

while interfering reflections of the sound would be played from speakers located in different directions.

8.2.2 Cocktail party effect

The importance of our work in furthering our understanding of the cocktail party effect comes from the fact that spatial cues are efficient cues to group speech segment (Bronkhorst, 2015). The fact that the auditory system is capable of extracting relevant information even in the presence of competing noise ensures that reliable spatial cues can be used during a cocktail party situation. For instance, Loizou et al. (2009) showed that bilateral CI users were worse at performing spatial release from masking in a cocktail party situation than normal hearing listeners. This result is possibly due to the absence of precedence effect behaviours in CI users. Adding an adaptation stage to cochlear implants and therefore allowing the precedence effect to take place is promising to improve spatial release from masking. This assumption could be tested in a similar way as Loizou et al. (2009), by asking subjects to repeat as many words as possible from a target location while multiple speech interferers are played simultaneously from different directions.

8.2.3 Multiple sound source localization and tracking

More generally, the mechanisms described could be used to work on improving multiple sound source localization and tracking. By using a wide range of stimuli and by integrating the different mechanisms described in chapter 4 and 5 to a sound localization algorithm similar to the one described in Goodman and Brette (2010), we could evaluate the performance of our models on localizing and tracking multiple sounds and compare it to the performance of the algorithm without the mechanisms implemented. Positive results could lead to interesting applications in the field of robotics.

8.2.4 Interaural level difference extraction in sound onsets

The core of the work presented here focused on the computation of interaural time differences in reverberant environments. One of the motivations behind studying ITD computation is the finding that the precedence effect is stronger for sounds lateralized with interaural time differences (Zurek, 1980; Krumbholz and Nobbe, 2002; Brown and Stecker, 2013; Brown et al., 2015). However, onset dominance was still observed in free-field conditions when sounds were lateralized by ILD (Stecker and Hafter, 2002; Stecker et al., 2013). Whether or not the mechanisms involved in the onset extraction of ILDs are similar to the ones proposed for extracting onset ITD is not yet understood. Reed and Blum (1990) proposed a model for ILD computation in the LSO. In this model, LSO cells are organized serially by the value of their firing threshold. Spikes coming from one side of the head provide excitatory inputs for the LSO cells while spikes coming from the opposite side of the head provide inhibitory inputs to the neurons. The LSO cells consequently compute the difference in spiking activity produced by each side of the head. In this serial threshold arrangement, the first cell that does not fire corresponds to the cell whose inputs cancel each other completely and therefore codes for the ILD value. On top of this architecture, it would be possible to build an additional layer capable of performing the same type of inhibitory computations as the ones described in chapter 3 and 5, and study whether the network could extract ILD cues in sound onsets. Abstract adaptation and onset behaviour models as described in chapters 4 and 5 are also likely to be useful tools in understanding ILDs computation better

8.2.5 The role of monaural spectral cues in the precedence effect

The precedence effect is often classified as a binaural perceptual phenomenon due to the fact that it is mostly studied with sounds lateralized by either ITD or ILD. While some aspects of the precedence effect seems to be reproducible in monaural conditions (Rakerd and Hartmann, 2005), it is believed that localization dominance cannot be achieved with monaural cues only (Dizon and Litovsky, 2004; Macpherson and Wagner, 2008). Designing techniques capable of accounting for localization dominance in monaural listening conditions would certainly be of

interest and would have immediate applications in hearing aid technologies.

8.2.6 The precedence effect with available visual cues

The precedence effect is observable without any visual information available to the listener. The influence of visual cues on localization dominance is still quite unknown. Available findings show that increasing visual attention to the first arriving sound improves the percept of its location (Bishop et al., 2011; Brown et al., 2015) compared to level differences. The effect of visual cues on the precedence effect could be investigated in a similar manner to a study we recently conducted (Steadman et al., 2019), through the use of virtual audio systems combined with virtual reality headsets.

8.2.7 Canonical mechanisms

We showed in chapters 4 and 5 that generic neural circuits could account for the data by Dietz et al. (2013) and Hu et al. (2017). This approach, along the lines of O'Leary et al. (2015), is likely to be fruitful in tackling different challenges both in the fi eld of auditory research and neuroscience in general.

8.3 Summary

We found that across multiple stages of the auditory pathway, several mechanisms (adaptation, onset firing and lateral inhibition), that are not specific to the auditory system, could lead to the extraction of temporal cues in the rising portion of sounds, in line with current findings in the field. This result opens us to the possibility that the computational functions carried out by neural circuits in the auditory pathway are not specialized to perform sound localization in reverberant environments but instead may be the results of evolutionnary exaptation.

Bibliography

Akeroyd, M. (2017). A binaural cross-correlogram toolbox for matlab.

Azadpour, M. and R. L. Smith (2016). Enhancing speech envelope by integrating hair-cell adaptation into cochlear implant processing. *Hearing research 342*, 48–57.

Bacon, S. P., R. R. Fay, and A. N. Popper (2004). *Compression: from cochlea to cochlear implants*. Springer.

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.

Bernstein, L. R. and C. Trahiotis (1994). Detection of interaural delay in high-frequency sinusoidally amplitude-modulated tones, two-tone complexes, and bands of noise. *The Journal of the Acoustical Society of America 95*(6), 3561–3567.

Bernstein, L. R. and C. Trahiotis (1996). On the use of the normalized correlation as an index of interaural envelope correlation. *The Journal of the Acoustical Society of America 100*(3), 1754–1763.

Bernstein, L. R. and C. Trahiotis (2002). Enhancing sensitivity to interaural delays at high frequencies by using transposed stimuli. *The Journal of the Acoustical Society of America 112*(3), 1026–1036.

Bishop, C. W., S. London, and L. M. Miller (2011). Visual influences on echo suppression. *Current Biology 21*(3), 221–225.

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.

Brand, A., O. Behrend, T. Marquardt, D. McAlpine, and B. Grothe (2002). Precise inhibition is essential for microsecond interaural time difference coding. *Nature 417*(6888), 543.

Breebaart, J., S. van de Par, and A. Kohlrausch (2001). Binaural processing model based on contralateral inhibition. ii. dependence on spectral parameters. *The Journal of the Acoustical Society of America 110*(2), 1089–1104.

Bregman, A. (1990). Auditory scene analysis: The perceptual organization of sound. cambridge, ma, us.

Bronkhorst, A. W. (2015). The cocktail-party problem revisited: Early processing and selection of multi-talker speech. *Attention, Perception, & Psychophysics 77*(5), 1465–1487.

Brown, A. and G. Stecker (2013). The precedence effect in sound localization: fusion and lateralization measures for pairs and trains of clicks lateralized by interaural time and level differences. *J Acoust Soc Am 133*, 2883–2898.

Brown, A. D., H. G. Jones, A. Kan, T. Thakkar, G. C. Stecker, M. J. Goupell, and R. Y. Litovsky (2015). Evidence for a neural source of the precedence effect in sound localization. *Journal of neurophysiology 114*(5), 2991–3001.

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.

Bruce, I. C., Y. Erfani, and M. S. Zilany (2018). A phenomenological model of the synapse between the inner hair cell and auditory nerve: Implications of limited neurotransmitter release sites. *Hearing research 360*, 40–54.

Carney, L. H., T. Li, and J. M. McDonough (2015). Speech coding in the brain: representation of vowel formants by midbrain neurons tuned to sound fluctuations. *Eneuro 2*(4).

Carr, C. E. and M. Konishi (1988). Axonal delay lines for time measurement in the owl's brainstem. *Proceedings of the National Academy of Sciences 85*(21), 8311–8315.

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

Cherry, E. C. (1953b). Some experiments on the recognition of speech, with one and with two ears. *The Journal of the acoustical society of America 25*(5), 975–979.

Clarey, J. C., P. Barone, W. A. Irons, F. K. Samson, and T. J. Imig (1995). Comparison of noise and tone azimuth tuning of neurons in cat primary auditory cortex and medical geniculate body. *Journal of neurophysiology 74*(3), 961–980.

Clopath, C., A. Longtin, and W. Gerstner (2008). An online hebbian learning rule that performs independent component analysis. *BMC Neuroscience 9*(1), O13.

Colburn, H. S. (1973). Theory of binaural interaction based on auditory-nerve data. i. general strategy and preliminary results on interaural discrimination. *The Journal of the Acoustical Society of America 54*(6), 1458–1470.

Cooper, N. and W. Rhode (1997). Mechanical responses to two-tone distortion products in the apical and basal turns of the mammalian cochlea. *Journal of neurophysiology 78*(1), 261–270.

Darrow, K. N., M. C. Slama, E. D. Kozin, M. Owoc, K. Hancock, J. Kempfle, A. Edge, S. Lacour, E. Boyden, D. Polley, et al. (2015). Optogenetic stimulation of the cochlear nucleus using channelrhodopsin-2 evokes activity in the central auditory pathways. *brain research 1599*, 44–56.

Darwin, C. J. (2005). Pitch and auditory grouping. In *Pitch*, pp. 278–305. Springer.

Darwin, C. J., D. S. Brungart, and B. D. Simpson (2003). Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers. *The Journal of the Acoustical Society of America 114*(5), 2913–2922.

Dau, T., B. Kollmeier, and A. Kohlrausch (1997a). Modeling auditory processing of amplitude modulation. i. detection and masking with narrow-band carriers. *The Journal of the Acoustical Society of America 102*(5), 2892–2905.

Dau, T., B. Kollmeier, and A. Kohlrausch (1997b). Modeling auditory processing of amplitude modulation. ii. spectral and temporal integration. *The Journal of the Acoustical Society of America 102*(5), 2906–2919.

Dau, T., D. Püschel, and A. Kohlrausch (1996a). A quantitative model of the effectivesignal processing in the auditory system. i. model structure. *The Journal of the Acoustical Society of America 99*(6), 3615–3622.

Dau, T., D. Püschel, and A. Kohlrausch (1996b). A quantitative model of the effectivesignal processing in the auditory system. ii. simulations and measurements. *The Journal of the Acoustical Society of America 99*(6), 3623–3631.

Devore, S., A. Ihlefeld, K. Hancock, B. Shinn-Cunningham, and B. Delgutte (2009, apr). Accurate Sound Localization in Reverberant Environments Is Mediated by Robust Encoding of Spatial Cues in the Auditory Midbrain. *Neuron 62*(1), 123–134.

Dietz, M., M. Klein-Hennig, and V. Hohmann (2015). The influence of pause, attack, and decay duration of the ongoing envelope on sound lateralization. *The Journal of the Acoustical Society of America 137*(2), EL137–EL143.

Dietz, M., J.-H. Lestang, P. Majdak, R. M. Stern, T. Marquardt, S. D. Ewert, W. M. Hartmann, and D. F. Goodman (2018). A framework for testing and comparing binaural models. *Hearing research 360*, 92–106.

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 110*(37), 15151–15156.

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), 1973–1985.

Dietz, M., L. Wang, D. Greenberg, and D. McAlpine (2016). Sensitivity to interaural time differences conveyed in the stimulus envelope: estimating inputs of binaural neurons through the temporal analysis of spike trains. *Journal of the Association for Research in Otolaryngology 17*(4), 313–330.

Dizon, R. M. and R. Y. Litovsky (2004). Localization dominance in the median-sagittal plane: effect of stimulus duration. *The Journal of the Acoustical Society of America 115*(6), 3142–3155.

Eglen, S. J., B. Marwick, Y. O. Halchenko, M. Hanke, S. Sufi, P. Gleeson, R. A. Silver, A. P. Davison, L. Lanyon, M. Abrams, et al. (2017). Toward standard practices for sharing computer code and programs in neuroscience. *Nature neuroscience 20*(6), 770.

Ewert, S. D. (2013). Afca modular framework for running psychoacoustic experiments and computational perception models. In *Proceedings of the international conference on acoustics AIA-DAGA*, pp. 1326–1329.

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.

Fitzpatrick, D. C., S. Kuwada, D. Kim, K. Parham, and R. Batra (1999). Responses of neurons to click-pairs as simulated echoes: auditory nerve to auditory cortex. *The Journal of the Acoustical Society of America 106*(6), 3460–3472.

Fontaine, B., D. F. Goodman, V. Benichoux, and R. Brette (2011). Brian hears: online auditory processing using vectorization over channels. *frontiers in Neuroinformatics 5*, 9.

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.

Glasberg, B. R. and B. C. Moore (2000). Frequency selectivity as a function of level and frequency measured with uniformly exciting notched noise. *The Journal of the Acoustical Society of America 108*(5), 2318–2328.

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

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.

Goodman, D. F. and R. Brette (2010). Spike-timing-based computation in sound localization. *PLoS computational biology 6*(11), e1000993.

Grantham, D. W., D. H. Ashmead, T. A. Ricketts, D. S. Haynes, and R. F. Labadie (2008). Interaural time and level difference thresholds for acoustically presented signals in post-lingually deafened adults fitted with bilateral cochlear implants using cis+ processing. *Ear and hearing 29*(1), 33–44.

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

Hammershøi, D. and H. Møller (1996). Sound transmission to and within the human ear canal. *The Journal of the Acoustical Society of America 100*(1), 408–427.

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

Hartmann, W. M. (1983). Localization of sound in rooms. *The Journal of the Acoustical Society of America 74*(5), 1380–1391.

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.

Hawley, M. L., R. Y. Litovsky, and J. F. Culling (2004). The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer. *The Journal of the Acoustical Society of America 115*(2), 833–843.

Heffner, H. (1997). The role of macaque auditory cortex in sound localization. *Acta Oto-Laryngologica 117*(sup532), 22–27.

Hirsh, I. (1948). Binaural summation and interaural inhibition as a function of the level of masking noise. *The American journal of psychology 61*(2), 205–213.

Hohmann, V. (2002). Frequency analysis and synthesis using a gammatone filterbank. *Acta Acustica united with Acustica 88*(3), 433–442.

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.

Hyvärinen, A., J. Karhunen, and E. Oja (2001). Independent component analysis, adaptive and learning systems for signal processing, communications, and control. *John Wiley & Sons, Inc 1*, 11–14.

Ivarsson, C., Y. De Ribaupierre, and F. De Ribaupierre (1988). Influence of auditory localization cues on neuronal activity in the auditory thalamus of the cat. *Journal of neurophysiology 59*(2), 586–606.

Jeffress, L. A. (1948). A place theory of sound localization. *Journal of comparative and physiological psychology 41*(1), 35.

Jepsen, M. L., S. D. Ewert, and T. Dau (2008). A computational model of human auditory signal processing and perception. *The Journal of the Acoustical Society of America 124*(1), 422–438.

Johnson, D. H. and D. E. Dudgeon (1993). *Array signal processing: concepts and techniques*. PTR Prentice Hall Englewood Cliffs.

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*.

Joris, P. X. and T. C. Yin (1992). Responses to amplitude-modulated tones in the auditory nerve of the cat. *The Journal of the Acoustical Society of America 91*(1), 215–232.

Kandler, K., A. Clause, and J. Noh (2009). Tonotopic reorganization of developing auditory brainstem circuits. *Nature neuroscience 12*(6), 711.

Kelly, J. B. and P. W. Judge (1985). Effects of medial geniculate lesions on sound localization by the rat. *Journal of neurophysiology 53*(2), 361–372.

Kerber, I. S. and I. B. U. Seeber (2012). Sound localization in noise by normal-hearing listeners and cochlear implant users. *Ear and hearing 33*(4), 445.

Klein-Hennig, M., M. Dietz, V. Hohmann, and S. D. Ewert (2011). The influence of different segments of the ongoing envelope on sensitivity to interaural time delays. *The Journal of the Acoustical Society of America 129*(6), 3856–3872.

Konishi, M. (2003). Coding of auditory space. *Annual review of neuroscience 26*(1), 31–55.

Köppl, C. (1997). Phase locking to high frequencies in the auditory nerve and cochlear nucleus magnocellularis of the barn owl, tyto alba. *Journal of Neuroscience 17*(9), 3312–3321.

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

Krumbholz, K. and A. Nobbe (2002). Buildup and breakdown of echo suppression for stimuli presented over headphonesthe effects of interaural time and level differences. *The Journal of the Acoustical Society of America 112*(2), 654–663.

Leek, M. R. (2001). Adaptive procedures in psychophysical research. *Perception & psychophysics 63*(8), 1279–1292.

Lesica, N. A., A. Lingner, and B. Grothe (2010). Population coding of interaural time differences in gerbils and barn owls. *Journal of Neuroscience 30*(35), 11696–11702.

Lestang, J.-H. and D. F. Goodman (2017). The roles of inhibition and adaptation for spatial hearing in difficult listening conditions. *The Journal of the Acoustical Society of America 141*(5), 3637–3637.

Lestang, J.-H. and D. F. Goodman (2019). Canonical brain computations account for perceived sound source location. *bioRxiv*, 687178.

Licklider, J. (1948). The influence of interaural phase relations upon the masking of speech by white noise. *The Journal of the Acoustical Society of America 20*(2), 150–159.

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.

Litovsky, R. Y. and B. G. Shinn-Cunningham (2001). Investigation of the relationship among three common measures of precedence: Fusion, localization dominance, and discrimination suppression. *The Journal of the Acoustical Society of America 109*(1), 346–358.

Loizou, P. C., Y. Hu, R. Litovsky, G. Yu, R. Peters, J. Lake, and P. Roland (2009). Speech recognition by bilateral cochlear implant users in a cocktail-party setting. *The Journal of the Acoustical Society of America 125*(1), 372–383.

Lorenzi, C., S. Gatehouse, and C. Lever (1999). Sound localization in noise in hearing-impaired listeners. *The Journal of the Acoustical Society of America 105*(6), 3454–3463.

Lorenzi, C., C. Micheyl, and F. Berthommier (1995). Neuronal correlates of perceptual amplitude-modulation detection. *Hearing research 90*(1-2), 219–227.

Macpherson, E. and M. Wagner (2008). Temporal weighting of cues for vertical-plane sound localization. In *Assoc. Res. Otolaryngol. Abstr*, Volume 31, pp. 301.

Macpherson, E. A. and J. C. Middlebrooks (2002). Listener weighting of cues for lateral angle: the duplex theory of sound localization revisited. *The Journal of the Acoustical Society of America 111*(5), 2219–2236.

Malhotra, S. and S. G. Lomber (2007). Sound localization during homotopic and heterotopic bilateral cooling deactivation of primary and nonprimary auditory cortical areas in the cat. *Journal of Neurophysiology 97*(1), 26–43.

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

McAlpine, D. and B. Grothe (2003). Sound localization and delay lines–do mammals fit the model? *Trends in neurosciences 26*(7), 347–350.

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*.

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

Meddis, R., M. J. Hewitt, and T. M. Shackleton (1990). Implementation details of a computation model of the inner hair-cell auditory-nerve synapse. *The Journal of the Acoustical Society of America 87*(4), 1813–1816.

Meddis, R., E. A. Lopez-Poveda, R. R. Fay, and A. N. Popper (2010). *Computational models of the auditory system*, Volume 35. Springer.

Middlebrooks, J. C. (1992). Narrow-band sound localization related to external ear acoustics. *The Journal of the Acoustical Society of America 92*(5), 2607–2624.

Middlebrooks, J. C. and J. D. Pettigrew (1981). Functional classes of neurons in primary auditory cortex of the cat distinguished by sensitivity to sound location. *Journal of Neuroscience 1*(1), 107–120.

Moore, B. C. and B. R. Glasberg (1983). Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. *The journal of the acoustical society of America 74*(3), 750–753.

Moore, B. C. and B. R. Glasberg (1987). Formulae describing frequency selectivity as a function of frequency and level, and their use in calculating excitation patterns. *Hearing research 28*(2-3), 209–225.

Moore, B. C., R. W. Peters, and B. R. Glasberg (1990). Auditory filter shapes at low center frequencies. *The Journal of the Acoustical Society of America 88*(1), 132–140.

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.

Patterson, R. D. (1976). Auditory filter shapes derived with noise stimuli. *The Journal of the Acoustical Society of America 59*(3), 640–654.

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.

Rakerd, B. and W. M. Hartmann (2005). Localization of noise in a reverberant environment. In *Auditory Signal Processing*, pp. 413–421. Springer.

Rayleigh, L. (1875). On our perception of the direction of a source of sound. *Proceedings of the Musical Association 2*, 75–84.

Reed, M. C. and J. J. Blum (1990). A model for the computation and encoding of azimuthal information by the lateral superior olive. *The Journal of the Acoustical Society of America 88*(3), 1442–1453.

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

Rhode, W. S. (1971). Observations of the vibration of the basilar membrane in squirrel monkeys using the mössbauer technique. *The Journal of the Acoustical Society of America 49*(4B), 1218–1231.

Rice, J. J., B. J. May, G. A. Spirou, and E. D. Young (1992). Pinna-based spectral cues for sound localization in cat. *Hearing research 58*(2), 132–152.

Roffler, S. K. and R. A. Butler (1968). Factors that influence the localization of sound in the vertical plane. *The Journal of the Acoustical Society of America 43*(6), 1255–1259.

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.

Ruggero, M. A., N. C. Rich, L. Robles, and B. G. Shivapuja (1990). Middle-ear response in the chinchilla and its relationship to mechanics at the base of the cochlea. *The Journal of the Acoustical Society of America 87*(4), 1612–1629.

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.

Saenz, M. and D. R. Langers (2014). Tonotopic mapping of human auditory cortex. *Hearing research 307*, 42–52.

Savin, C., P. Joshi, and J. Triesch (2010). Independent component analysis in spiking neurons. *PLoS computational biology 6*(4), e1000757.

Shinn-Cunningham, B. and K. Kawakyu (2003). Neural representation of source direction in reverberant space. In *2003 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (IEEE Cat. No. 03TH8684)*, pp. 79–82. IEEE.

Søndergaard, P. L. and P. Majdak (2013). The auditory modeling toolbox. In *The technology of binaural listening*, pp. 33–56. Springer.

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.

Steadman, M. A., C. Kim, J.-H. Lestang, D. F. Goodman, and L. Picinali (2019). Short-term effects of sound localization training in virtual reality. *bioRxiv*, 207753.

Stecker, G. C. and E. R. Hafter (2002). Temporal weighting in sound localization. *The Journal of the Acoustical Society of America 112*(3), 1046–1057.

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.

Stecker, G. C., J. D. Ostreicher, and A. D. Brown (2013). Temporal weighting functions for interaural time and level differences. iii. temporal weighting for lateral position judgments. *The Journal of the Acoustical Society of America 134*(2), 1242–1252.

Stimberg, M., D. F. M. Goodman, V. Benichoux, and R. Brette (2014). Equation-oriented specification of neural models for simulations. *Frontiers in Neuroinformatics 8*, 6.

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

van de Par, S. and A. Kohlrausch (1999). Dependence of binaural masking level differences on center frequency, masker bandwidth, and interaural parameters. *The Journal of the Acoustical Society of America 106*(4), 1940–1947.

Verhulst, S., A. Altoe, and V. Vasilkov (2018). Computational modeling of the human auditory periphery: Auditory-nerve responses, evoked potentials and hearing loss. *Hearing research 360*, 55–75.

Verschooten, E., S. Shamma, A. J. Oxenham, B. C. Moore, P. X. Joris, M. G. Heinz, and C. J. Plack (2019). The upper frequency limit for the use of phase locking to code temporal fine structure in humans: A compilation of viewpoints. *Hearing research 377*, 109–121.

Von Békésy, G. and E. G. Wever (1960). *Experiments in hearing*, Volume 8. McGraw-Hill New York.

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.

Wang, L. and H. S. Colburn (2012). A modeling study of the responses of the lateral superior olive to ipsilateral sinusoidally amplitude-modulated tones. *Journal of the Association for Research in Otolaryngology 13*(2), 249–267.

Weber, D. L. (1977). Growth of masking and the auditory filter. *The Journal of the Acoustical Society of America 62*(2), 424–429.

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.

Whitfield, I., J. Cranford, R. Ravizza, and I. Diamond (1972). Effects of unilateral ablation of auditory cortex in cat on complex sound localization. *Journal of Neurophysiology 35*(5), 718–731.

Wightman, F. L. and D. J. Kistler (1997). Monaural sound localization revisited. *The Journal of the Acoustical Society of America 101*(2), 1050–1063.

Wrobel, C., A. Dieter, A. Huet, D. Keppeler, C. J. Duque-Afonso, C. Vogl, G. Hoch, M. Jeschke, and T. Moser (2018). Optogenetic stimulation of cochlear neurons activates the auditory pathway and restores auditory-driven behavior in deaf adult gerbils. *Science translational medicine 10*(449), eaao0540.

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.

Yin, T. (1994). Physiological correlates of the precedence effect and summing localization in the inferior colliculus of the cat. *Journal of Neuroscience 14*(9), 5170–5186.

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), 1540–1553.

Zurek, P. M. (1980). The precedence effect and its possible role in the avoidance of interaural ambiguities. *The Journal of the Acoustical Society of America 67*(3), 952–964.

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

Appendices

Implementation Meddis model

Implementation of the Meddis model in Python, following the recommendations from Meddis et al. (1990).

```
from brian2 import *
from brian2.hears import *

def meddis(gfb,spikes='off',lval=2580,yval=5.05,rval=6580)
    eqsMeddis=Equations('''
        A=5:1
        B=300:1
        g=2000:1
        y:5.05:1
        l:25801
        r:65801
        x=66.31:1
        V:1
        k=g*((V+A)/(V+A+B)*1.0)*((V+A)*1.0>0):1
        dq/dt=(y*(1-q)+x*w-k*q)/second:1
        dc/dt=(k*q-l*c-r*c)/second:1
        dw/dt=(r*c-x*w)/second:1
        out=c:1
        ''')

        G=FilterbankGroup(gfb , 'V', eqsMeddis)

        G.c=G.k*G.y*1.0/(G.y*(G.l+G.r)+G.k*G.l)
        G.q=G.c*(G.l+G.r)*1.0/G.k
        G.w=G.c*G.r*1.0/G.x

return G
```

Implementation adaptation loops

This section shows the python implementation of the adaptation loops (Dau et al., 1996a,b) following the Matlab implementation by Stephan Ewert that can be found at https://github.com/model-initiative/model_initiative/blob/master/model_initiative/pathway_models/preprocessing/Adaptation%20Loops%20Dau%201996/nlal_lim_var.m.

```
import numpy as np

def dau_loops(nbrLoops, source, tau):

    b0 = np.zeros((nbrLoops))
    a1 = np.zeros((nbrLoops))
    tmp2 = np.zeros((nbrLoops))
    minLev = 0.00001

    for i in xrange(nbrLoops):
        b0[i] = 1.0 / (tau[i] * 44100)
        a1[i] = exp(-b0[i])
        b0[i] = 1 - a1[i]
        tmp2[i] = minLev**(0.5**(i + 1))
    minl = tmp2
    corr=tmp2[-1]
    mult=100./(1-corr)
    arr = np.array(source)
    arrout = np.array(source)
    limit = 10

    for i in xrange(source.nsamples):

        tmp1 = arr[i, 0]
        if tmp1 < minLev:
            tmp1 = minLev
        for j in xrange(nbrLoops):
            tmp1 = tmp1 / tmp2[j]
            maxvalue = (1 - minl[j]**2) * limit - 1
            factor = maxvalue * 2
            expfac = -2.0 / maxvalue
            offset = maxvalue - 1
            if tmp1 > 1:
                tmp1 = factor / (1.0 + exp(expfac * (tmp1 - 1))) - offset

            tmp2[j] = a1[j] * tmp2[j] + b0[j] * tmp1
            arrout[i, 0] = tmp1
    return arrout
```

Detailed onset model

The code below is the model of onset cells used in section .

```
function [spOut, vOut] = BC_onsettype_modelAWC(spEx, DT, Iext)
% %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% original: Adjusted Wang-Colburn Model of Lateral Superior Olive
% now: feed with parameters from Rothman-Manis (2003)to behave like
% onset-type cochlea nucleus neuron
% %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Input
%  spEx : excitatory input spike vector (number of inputs at each time step)
%  DT   : size of time step [ms]
%  Iext (optional) : external current [pA] (will be converted into uA)
%
% Output
%  spOut : number of output spikes at each time step
%  vOut  : membrane potential at each time step
% %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Notes on the input/output arguments
% + Iext is an optional input argument to specify (time-varying) external
%   input current. If this argument is not given (or given as an empty
%   vector), then the external current is assumed to be zero.
% + If the external current Iext is a scalar, then the model is assumed
%   to receive a constant current of size Iext [pA].
% + Total simulation time length is determined as below:
%  - If Iext is not given, given as an empty vector, or given as a scalar,
%    then the shorter of spEx and spIn determines the length.
%  - If Iext is a non-empty vectors, then the shortest of spEx, spIn, and
%    Iext determines the length.
% %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Units to use in the code
% (all parameters have to be converted into these units before use)
```

```
%  T  [ms]
%  I  [uA]
%  V  [mV]
%  g  [mS]
%  C  [uF]
% %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Notes on the model
%  The model is based on the model presented in Wang & Colburn (2012),
%  but with modifications of the parameters to replicate the bushy cell
%  onset-type described by Rothman & Manis (2003). Modifications include:
%  membrane characteristics were adjusted to parameters shown in Table 1
%  (Rothman & Manis, June 2003), inhibitory input was removed,
%  hyperpolarization-activated cation current was added according
%  to the equations of the Rothman-Manis (2003, equations A24-A26).
% %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% References
% Wang L, Colburn HS (2012) JARO 13: 249-267
%  "A modeling study of the responses of the lateral superior olive
%   to ipsilateral sinusoidally amplitude-modulated tones"
% Rothman JS, Manis PB (2003) J Neurophysiol 89: 3097  3113
%  "The roles potassium currents play in regulating the electrical
%   activity of ventral cochlear nucleus neurons"
% Ashida G, Tollin DJ, Kretzberg J (2017) PLoS Comput Biol 13(12): e1005903.
% "Physiological models of the lateral superior olive"
% https://doi.org/10.1371/journal.pcbi.1005903
% %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Revisions
% Created (ver 0.9): May 16, 2017 by Go Ashida
% Updated references info (ver 1.0): Aug 02, 2017 by Lisa Schmors
% %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% If you find a bug, please report to GA at go.ashida@uni-oldenburg.de
% %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

```
%%%%% Copyright 2017 Go Ashida (go.ashida@uni-oldenburg.de) %%%%%%%%%%%%
% Permission is hereby granted under the Apache License, Version 2.0;
% Users of this file must be in compliance with this license, a copy of
% which may be obtained at http://www.apache.org/licenses/LICENSE-2.0
% This file is provided on an "AS IS" basis, WITHOUT WARRANTIES OR
% CONDITIONS OF ANY KIND, either express or implied.
% %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%% simulation parameters
% If Iext is not given, then create it as an empty vector
if ~exist('Iext','var')
 Iext = [];
end

% Determine the simulation time length
if length(Iext)>1
 Nsteps = min([length(spEx), length(Iext)]);
else
 Nsteps = length(spEx);
end

% Make external input vector depending on the input argument
if length(Iext)>1 % if the Iext is a vector, convert it from [pA] to [uA]
 jext = Iext * 1e-6;
elseif length(Iext)==1 % if Iext is a scalar, then set a constant input
 jext = ones(1,Nsteps) * Iext(1) * 1e-6;
else % if Iext is an empty vector, then use zero input
 jext = zeros(1,Nsteps);
end

% Create output vectors
spOut = zeros(1,Nsteps);
vOut = zeros(1,Nsteps);

%% model parameters

% membrane parameters
Cs  = 12e-6;          % 312*1e-6[uF] = 12[pF]
gLL = 2e-6;           % 2*1e-6[mS] = 2[nS]
```

```
102 gKL = 200e-6;          % 200*1e-6[mS] = 200[nS]
103 gKH = 150e-6;          % 150*1e-6[mS] = 150[nS]
104 gNa = 1000e-6;         % 1000*1e-6[mS] = 1000[nS]
105 gH  = 20e-6;           % 20*1e-6[ms] = 20[nS] hyperpolarization
        conductance
106 EL = -65.0;            % [mV] leak reversal potential
107 EK = -70.0;            % [mV] potassium reversal potential
108 EN = +55.0;            % [mV] sodium reversal potenti
109 EH = -43;              % [mV] hyperpolarization reversal potential
110
111 % synaptic input parameters
112 Ge =  3.0;             % from Dietz et al. 2016; [nS] excitatory
        synaptic input amplitude
113 Trise = 0.1;           % from Dietz et al. 2016; [ms]
114 Tfall = 1;             % from Dietz et al. 2016; [ms]
115 Eex =  0.0; % [mV] excitatory synaptic input reversal potential
116
117 % temperature factors
118 Temp_body = 37.0;
119 Temp_KLVA = 22.0;
120 Temp_KHVA = 22.0;
121 Temp_Na   = 22.0;
122 Temp_H    = 22.0;
123 Q10_KL = 3.0;
124 Q10_KH = 3.0;
125 Q10_Na = 3.0;
126 Q10_H  = 3.0;
127 T_KL = ( Q10_KL ^ ((Temp_body-Temp_KLVA)/10.0) );
128 T_KH = ( Q10_KH ^ ((Temp_body-Temp_KHVA)/10.0) );
129 T_Na = ( Q10_Na ^ ((Temp_body-Temp_Na)  /10.0) );
130 T_H  = ( Q10_H  ^ ((Temp_body-Temp_H)   /10.0) );
131
132 % initialization
133 vRest = -64.9;                % from Dietz et al 2016 onset type CN
134 vOut(1) = vRest;
135 w = winf(vRest);
136 z = zinf(vRest);
137 n = ninf(vRest);
138 p = pinf(vRest);
139 m = minf(vRest);
140 h = hinf(vRest);
141 r = rinf(vRest);
142
143 % spike detection thresholds
144 vTh1 = -30.0; % [mV] spike initiation
145 vTh2 = -45.0; % [mV] spike end
146 spikeflag = 0;  % flag for spiking
147
```

```
148 %% synaptic inputs
149 gEx = BetaSynapse(spEx, DT, Ge, Trise, Tfall); % [nS]
150
151 %% main loop
152 for t=1:Nsteps-1
153
154  % calculating all values at time t
155  u = vOut(t); % potential
156  iLL = gLL * (EL-u); % leak current
157  iKL = gKL * (EK-u) * ( w^4 * z ); % KLVA current
158  iKH = gKH * (EK-u) * ( 0.85 * n * n + 0.15 * p ); % KHVA current
159  iNa = gNa * (EN-u) * ( m^3 * h ); % Na current
160  iH  = gH * (EH-u) * r;
161
162  iEx = gEx(t) * (Eex-u) * 1e-6; % [uA] excitatory synaptic current
163  itot = iLL + iKL + iKH + iNa + iEx + iH + jext(t); % total current
164
165  % step forward
166  v = vOut(t) + (itot/Cs) * DT;
167  w = w + T_KL * ( winf(u) - w )/ tauw(u) * DT;
168  z = z + T_KL * ( zinf(u) - z )/ tauz(u) * DT;
169  n = n + T_KH * ( ninf(u) - n )/ taun(u) * DT;
170  p = p + T_KH * ( pinf(u) - p )/ taup(u) * DT;
171  m = m + T_Na * ( minf(u) - m )/ taum(u) * DT;
172  h = h + T_Na * ( hinf(u) - h )/ tauh(u) * DT;
173  r = r + T_H  * ( rinf(u) - r )/ taur(u)  * DT;
174
175  % check for spiking
176  if(spikeflag) % if spiking, then look for end of spike
177    spOut(t+1) = 0;
178    if(v<=vTh2)
179     spikeflag = 0; % reset the flag
180    end
181  else % if not spiking, then look for spike threshold crossing
182    if(v>=vTh1)
183     spOut(t+1) = 1; % report spike generation
184     spikeflag = 1; % set the flag
185    else
186     spOut(t+1) = 0;
187    end
188  end
189
190  % save membrane potential
191  vOut(t+1) = v;
192
193 end % end of main loop
194
195 %% end of main function
```

```
196 end % corresponding to the first line of this function declaration
197
198 %% internal functions for channel kinetics
199 function x = winf(v) % KLVA activation : steady state
200  x = 1.0 / sqrt(sqrt( 1.0 + exp( -(v+48.0)/6.0 ) ));
201 end
202 function x = tauw(v) % KLVA activation : time constant
203  x = 1.5 + 100.0 / ( 6.0*exp( (v+60.0)/6.0 ) + 16.0*exp( -(v+60.0)
       /45.0 ) );
204 end
205 function x = zinf(v) % KLVA inactivation : steady state
206  x = 0.5 + 0.5 / ( 1.0 + exp( (v+71.0)/10.0 ) );
207 end
208 function x = tauz(v) % KLVA inactivation : time constant
209  x = 50.0 + 1000.0 / ( exp( (v+60.0)/20.0 ) + exp( -(v+60.0)/8.0 ) )
       ;
210 end
211 function x = ninf(v) % KHVA activation 1 : steady state
212  x = 1.0 / sqrt( 1.0 + exp( -(v+15.0)/5.0 ) );
213 end
214 function x = taun(v) % KHVA activation 1 : time constant
215  x = 0.7 + 100.0 / ( 11.0*exp( (v+60.0)/24.0 ) + 21.0*exp( -(v+60.0)
       /23.0 ) );
216 end
217 function x = pinf(v) % KHVA activation 2 : steady state
218  x = 1.0 / ( 1.0 + exp( -(v+23.0)/6.0 ) );
219 end
220 function x = taup(v) % KHVA activation 2 : time constant
221  x = 5.0 + 100.0 / ( 4.0*exp( (v+60.0)/32.0 ) + 5.0*exp( -(v+60.0)
       /22.0 ) );
222 end
223 function x = minf(v) % Na activation : steady state
224  x = 1.0 / ( 1.0 + exp( -(v+38.0)/7.0 ) );
225 end
226 function x = taum(v) % Na activation : time constant
227  x = 0.04 + 10.0 / ( 5.0*exp( (v+60.0)/18.0 ) + 36.0*exp( -(v+60.0)
       /25.0) );
228 end
229 function x = hinf(v) % Na inactivation : steady state
230  x = 1.0 / ( 1.0 + exp( (v+65.0)/6.0 ) );
231 end
232 function x = tauh(v) % Na inactivation : time constant
233  x = 0.6 + 100.0 / ( 7.0*exp( (v+60.0)/11.0 ) + 10.0*exp( -(v+60.0)
       /25.0) );
234 end
235
236 function x = rinf(v) % Hyperpolarization-activation : steady state (
      see Rothman & Manis 2003; Equ. A24-26)
```

```
237 x = 1/(1+exp((v+76)/7));
238 end
239
240 function x = taur(v) % Hyperpolarization-activation : time constant
      (see Rothman & Manis 2003; Equ. A24-26)
241 x = 25 +1e5/(237* exp((v+60)/12)+17* exp(-(v+60)/14));
242 end
```

Apache license for detailed onset model

Apache License

Version 2.0, January 2004

http://www.apache.org/licenses/

TERMS AND CONDITIONS FOR USE, REPRODUCTION, AND DISTRIBUTION

1. Definitions.

"License" shall mean the terms and conditions for use, reproduction, and distribution as defined by Sections 1 through 9 of this document.

"Licensor" shall mean the copyright owner or entity authorized by the copyright owner that is granting the License.

"Legal Entity" shall mean the union of the acting entity and all other entities that control, are controlled by, or are under common control with that entity. For the purposes of this definition, "control" means (i) the power, direct or indirect, to cause the direction or management of such entity, whether by contract or otherwise, or (ii) ownership of fifty percent (50

"You" (or "Your") shall mean an individual or Legal Entity exercising permissions granted by this License.

"Source" form shall mean the preferred form for making modifications, including but not limited to software source code, documentation source, and configuration files.

"Object" form shall mean any form resulting from mechanical transformation or translation of a Source form, including but not limited to compiled object code, generated documentation, and conversions to other media types.

"Work" shall mean the work of authorship, whether in Source or Object form, made available under the License, as indicated by a copyright notice that is included in or attached to the work (an example is provided in the Appendix below).

"Derivative Works" shall mean any work, whether in Source or Object form, that is based on (or derived from) the Work and for which the editorial revisions, annotations, elaborations, or other modifications represent, as a whole, an original work of authorship. For the purposes of this License, Derivative Works shall not include works that remain separable from, or merely link (or bind by name) to the interfaces of, the Work and Derivative Works thereof.

"Contribution" shall mean any work of authorship, including the original version of the Work and any modifications or additions to that Work or Derivative Works thereof, that is intentionally submitted to Licensor for inclusion in the Work by the copyright owner or by an individual or Legal Entity authorized to submit on behalf of the copyright owner. For the purposes of this definition, "submitted" means any form of electronic, verbal, or written communication sent to the Licensor or its representatives, including but not limited to communication on electronic mailing lists, source code control systems, and issue tracking systems that are managed by, or on behalf of, the Licensor for the purpose of discussing and improving the Work, but excluding communication that is conspicuously marked or otherwise designated in writing by the copyright owner as "Not a Contribution."

"Contributor" shall mean Licensor and any individual or Legal Entity on behalf of whom a Contribution has been received by Licensor and subsequently incorporated within the Work.

2. Grant of Copyright License. Subject to the terms and conditions of this License, each Contributor hereby grants to You a perpetual, worldwide, non-exclusive, no-charge, royalty-free, irrevocable copyright license to reproduce, prepare Derivative Works of, publicly display,

publicly perform, sublicense, and distribute the Work and such Derivative Works in Source or Object form.

3. Grant of Patent License. Subject to the terms and conditions of this License, each Contributor hereby grants to You a perpetual, worldwide, non-exclusive, no-charge, royalty-free, irrevocable (except as stated in this section) patent license to make, have made, use, offer to sell, sell, import, and otherwise transfer the Work, where such license applies only to those patent claims licensable by such Contributor that are necessarily infringed by their Contribution(s) alone or by combination of their Contribution(s) with the Work to which such Contribution(s) was submitted. If You institute patent litigation against any entity (including a cross-claim or counterclaim in a lawsuit) alleging that the Work or a Contribution incorporated within the Work constitutes direct or contributory patent infringement, then any patent licenses granted to You under this License for that Work shall terminate as of the date such litigation is filed.

4. Redistribution. You may reproduce and distribute copies of the Work or Derivative Works thereof in any medium, with or without modifications, and in Source or Object form, provided that You meet the following conditions:

You must give any other recipients of the Work or Derivative Works a copy of this License; and You must cause any modified files to carry prominent notices stating that You changed the files; and You must retain, in the Source form of any Derivative Works that You distribute, all copyright, patent, trademark, and attribution notices from the Source form of the Work, excluding those notices that do not pertain to any part of the Derivative Works; and If the Work includes a "NOTICE" text file as part of its distribution, then any Derivative Works that You distribute must include a readable copy of the attribution notices contained within such NOTICE file, excluding those notices that do not pertain to any part of the Derivative Works, in at least one of the following places: within a NOTICE text file distributed as part of the Derivative Works; within the Source form or documentation, if provided along with the Derivative Works; or, within a display generated by the Derivative Works, if and wherever such third-party notices normally appear. The contents of the NOTICE file are for informational purposes only and do not modify the License. You may add Your own attribution notices within Derivative Works that You distribute, alongside or as an addendum to the NOTICE text from the Work, provided that such additional attribution notices cannot be construed as modifying the License.

You may add Your own copyright statement to Your modifications and may provide additional or different license terms and conditions for use, reproduction, or distribution of Your modifications, or for any such Derivative Works as a whole, provided Your use, reproduction, and distribution of the Work otherwise complies with the conditions stated in this License.

5. Submission of Contributions. Unless You explicitly state otherwise, any Contribution intentionally submitted for inclusion in the Work by You to the Licensor shall be under the terms and conditions of this License, without any additional terms or conditions. Notwithstanding the above, nothing herein shall supersede or modify the terms of any separate license agreement you may have executed with Licensor regarding such Contributions.

6. Trademarks. This License does not grant permission to use the trade names, trademarks, service marks, or product names of the Licensor, except as required for reasonable and customary use in describing the origin of the Work and reproducing the content of the NOTICE file.

7. Disclaimer of Warranty. Unless required by applicable law or agreed to in writing, Licensor provides the Work (and each Contributor provides its Contributions) on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied, including, without limitation, any warranties or conditions of TITLE, NON-INFRINGEMENT, MERCHANTABILITY, or FITNESS FOR A PARTICULAR PURPOSE. You are solely responsible for determining the appropriateness of using or redistributing the Work and assume any risks associated with Your exercise of permissions under this License.

8. Limitation of Liability. In no event and under no legal theory, whether in tort (including negligence), contract, or otherwise, unless required by applicable law (such as deliberate and

grossly negligent acts) or agreed to in writing, shall any Contributor be liable to You for damages, including any direct, indirect, special, incidental, or consequential damages of any character arising as a result of this License or out of the use or inability to use the Work (including but not limited to damages for loss of goodwill, work stoppage, computer failure or malfunction, or any and all other commercial damages or losses), even if such Contributor has been advised of the possibility of such damages.

9. Accepting Warranty or Additional Liability. While redistributing the Work or Derivative Works thereof, You may choose to offer, and charge a fee for, acceptance of support, warranty, indemnity, or other liability obligations and/or rights consistent with this License. However, in accepting such obligations, You may act only on Your own behalf and on Your sole responsibility, not on behalf of any other Contributor, and only if You agree to indemnify, defend, and hold each Contributor harmless for any liability incurred by, or claims asserted against, such Contributor by reason of your accepting any such warranty or additional liability.

END OF TERMS AND CONDITIONS