Mesoscopic dynamics of pitch processing in human auditory cortex

Author: Alejandro Tabas
First supervisor: Emili Balaguer-Ballester
Second supervisor: André Rupp
Third supervisor: Hamid Bouchachia

Thesis submitted in partial fullfilment of the requirements of Bournemouth University for the degree of

Doctor of Philosophy

February, 2017

Abstract

Pitch is a perceptual correlate of sound periodicity elicited by vibrating bodies; it plays a crucial role in music and speech. Although perceptual phenomenology of pitch has been studied for centuries, a detailed understanding of its underlying neural mechanisms is still lacking. Early theories suggesting that pitch is decoded in the peripheral auditory system fail to explain the perception of complex stimuli. More recent mechanistic models, focused on how subcortical structures process periodic discharges of the auditory nerve activity, are unable to explain fully key aspects of the processing dynamics observed during electrophysiological recordings. In this thesis, we propose a novel theory describing how subcortical representations of pitch-related information are integrated in cortex and how this integratory process gives rise to the dynamics observed in magnetoencephalographic (MEG) experimental recordings.

Auditory evoked fields recorded with MEG reveal a systematic deflection around 100 ms after stimulus' onset known as the N100m. This deflection consists of several components reflecting the onset of different perceptual dimensions of auditory stimuli such as pitch, timbre and loudness. The exact latency of the component elicited by pitch onset, known as the pitch onset response (POR), shows a strong linear relationship with the pitch of the stimulus. Our theory links the POR latency with processing time and explains, in a quantitative manner, the substrate of the relationship between processing time and pitch.

Cortical integration is described using a model of neural ensembles located in two adjacent areas, putatively located along the lateral portion of Heschl's Gyrus in human auditory cortex. Cortical areas are hierarchically structured and communicate with each other in a top-down fashion. Pitch processing is modelled as a multi-attractor system whose dynamics are driven by subcortical input. After tone onset, the system evolves from an initial equilibrium position to a new equilibrium state that represents the pitch elicited by the tone. A computational implementation of the model shows that: 1) the transient dynamics between equilibrium points explains the POR; 2) the latency of the transient is directly linked with the time required to reach the new equilibrium state; and 3) that such processing time depends linearly on the pitch of the stimuli.

Our theory also addresses the problem of how tones with several simultaneous pitch values are processed in cortex. In Western music, dyads comprising tones with different pitch values are judged as more consonant or more dissonant depending on the ratio of the periods of the involved sounds. The latency of the POR evoked by such dyads also presents a strong correlation with the perceived consonance: dissonant dyads generate later PORs than consonant dyads. Our theory of pitch processing describes consonance (dissonance) as a direct effect of harmonic collaboration (competition) during the cortical integration process: the cortical model shows that harmonic collaboration facilitates convergence, explaining why dissonant dyads require longer processing times and evoke later PORs than consonant dyads.

Contents

List of Figures

Acknowledgments

PhDs[1] are often described as taxing, obnoxious and exhausting torments; this is not only a PhD comics observation, but a rather universal belief, whose reach can be measured by the number of times people apologise after accidentally asking about the forbidden t-word. What is then, I often find myself asking to the other myself, what has made this adventure so marvellously fulfilling, so wholly enthralling, so fully stimulating?

Transmuting a potentially tedious three-years-and-a-half misery into the enlightening jubilant journey that I have experienced is far from being explainable in a few lines. Highly non-linear complex effects and a large number of variables had to collude so that I could write the last lines of this thesis without losing my marbles. Fortunately, all these effects and variables have names (and even surnames!), and the spaghetti monster of the PhD theses invented a special section to express my gratitude, acknowledging their crucial contribution to my sanity, and therefore to the completion this work.

Emili and Andre have overly exceeded their academic obligations in their generosity and dedication. Emili's brilliant scientific intuition, prominent positivity, skilful diplomacy, and restless commitment were everything a PhD student of my nature could possibly ask for. He has guided, during these three rainy years, not only this work, but also my scientific and personal development (and that was not included in the contract!); so I am happy and proud to call him not only my supervisor, but also my mentor and my friend.

Andre received me in his amazing lab in Heidelberg with open arms, a photographic lab, important lessonx in German gastronomy, and the best beer and espresso that Baden-Württemberg can offer. With great patience, he systematically brought me to his office, and shared with me his vast knowledge on experimental (and even theoretical) auditory neuroscience, showing me everything I know about MEG, the FFR, and Patterson's AIM (the reader should note that it would have been much easier to email me some papers and a few slides and wish me luck). His infinite generosity has granted us all the data and unpublished experimental results that any modellist could wish, and his extraordinary adaptability allowed us to focus on the research rather than in the rush when the time was tight and the computations slow.

The rare openness of key people at the Bournemouth University is worth to mention; specially the the affability and readiness of Naomi Bailey, who had already allocated me a group of friends even before I put a foot in town; and the availability and sense of humour of Malcolm Green, whose flexibility allowed me to stretch our conference budget beyond the imaginable (and who, I am sure, is still worried about the location of this stampcrab's shoes).

Amazing and loving people populate Andre's MEG lab in the Biomagnetism section of the Heidelberg University. For instance, Britta Kretzmar, who taught me the importance of the Katz and introduced me to the subject of pain (in both, humans and mice); Anna-Maria Hassel, who repeatedly welcomed me in the lab, supervised many of my crazy recordings, and (most importantly) showed me the golden path to the coffee machine; Barbara Burghardt, who patiently (and repeatedly) explained me the behaviour of the MEG monster in a mixture of Deutsch, English and (mostly) sign language; Helmut Riedel, who not only trained me in akkuStim, but also modified it so it could hold my kooky experimental paradigms; or Heide Rogatzki, who virtually guaranteed my viability by finding me a suitable place to live and lively fighting against the bureaucracy of the Heidelberg University in my behalf.

[1] Beware: Andalusian blood runs on my veins~~, so~~ and my brains are prune to exaggerations and oversentimentalisms; serious readers: feel free to skip the bah-humbug and jump directly to the next page :)

An important mention should be made to both of my parents, who supported me in the most orthogonal fashion a two-dimensional space can hold: my father, who funded my predoctoral education, no matter the absurdity of starting theoretical physics when what you really want to do is to work in the film industry; and my mother, who always helped me to find the silverlinings of my own decisions, pushed me to take the riskier routes, and virtually took care of 90% of my non-research-related life during the last weeks of my Thesis writing when I was supposed to be at home on holidays.

Following the family line, I have to mention my aunt Maribel for her perpetual (and unjustified) love, amazing culinary creations, and (together with my uncle Juan) for lending me her amazing house at the beach, witness of the writing of the first four chapters of this thesis.

Here I feel I should also make a place to Castro, the great Cuban dictator, for his enthusiastic generosity, superb empanadas, constant taxiing service, precious photographic support and, most importantly, for his haircut support.

Of extreme importance for the preservation of my sanity were my support groups in Bournemouth and Madrid. Bournemouth team was characterised by the specialisation of their members: my cousins Manolo and Cristina, my familiar, always present, always available, always loving, cornerstone; Rush, my exquisite intellectual guru, social enlightener, musical comrade, boardgame vanquisher, and planking master; Ali, my Persian other half, canoodling hairy crony and pipe smoking pal, and Naj, my hugging dispenser and saffron dealer; TO, my nightime accomplice, adventure allocator, aleologist in chief, and moonlight authority; Blondie, my constant agitator, skilled philosophical interlocutor, and hurricane cofighter; Nico, my hops instructor, physics raconteur, and favourite New Orleans clarinet consort; Ox, headmaster of the debate club, pure freshness of the bier connoisseurs, and endless source of laugh; Azahara, for her arty references, and her Spaniard time creations; Bastian, for sharing with me his extraordinary view of life; and, of course, Nan, my travelling associate, fellow foodie, espresso sharer, and chocolate supplier.

Madrid support group mainly encompasses family members: Pér, unconditional friend and everlasting companion; Karma, the mystic wizard, fountainhead of all creativeness of this world; Elena, the eldest of my comrades, literature concomitant, and beer coguzzler; Dolç, the terminal cinephile who, despite walking me over most phases of my life, managed to approximately love me in all of my shapes; Amiga, whose Satchmo's recordings warmed up the cold dark hours of the Madrilian night; Cam, whose home has been always open for the eclectic nonsense of the Hypatian nights, and secretly accelerates the bus's delay when she know I need her around; Amigo, who showed me the reach of relativism and still enlightens my walk through the path towards critical thinking; Jakob, master guru of my academic inspiration, and Nan, whose tender proximity and maternal love I still struggle to understand; Mats, who trained me to accenttchuate the positive, and keeps reinforcing it with her harmonic rationality once a year during the most traditional night of the Christmas season; Paloma, who generously supported me to leave Barcelona and tolerated my eccentricities for a myriad of lengthy years; and of course Alba, the last addition to the family, whose mellow distant lines, home-made olives, prompt hooves, quirky theatre shows, and Bunyanesque sleep-time generosity, greatly sweetened the onerous months of the thesis writing venture.

Mojácar,
Alemería,
23rd of October, 2016

Münstertal,
Breisgau-Hochschwarzwald,
25th of February, 2017

Author declaration

The author declares that this work was done mainly while in candidature for a PhD degree at the Bournemouth University. Where I have consulted the published work of others, this is always clearly attributed. Where I have quoted from the work of others, the source is always given.

With the exception of the measurements of the experimental data shown in Section 3.4.3, which was recorded by Anita Siebert and André Rupp; and the analysis and measurement of the experimental data shown in Section 5.2.2, and Section 5.2.3, which was recorded and analysed by Valeria Sebold, Martin Andermann, and André Rupp; this thesis is entirely my own work.

Results in Section 3.4 and a preliminary version of the results of Section 4.3.2 have been published before submission in [1] and [2], respectively.

Chapter 1

Introduction

Human sensory systems generate perceptual experiences by processing physical properties of the surrounding environment. For instance, the visual sensory system enables the human visual cortex to perceive brightness and colour, two perceptual entities respectively correlated with two physical properties of the incident electromagnetic radiation measured by the photoreceptors in the retina: amplitude and frequency [3].

Auditory entities like speech or music present complex structures that, like the components of visual stimuli, can be described as a conjunction of the physical properties of the auditory stimuli. Sound sensations correlate to the movement induced in the tympanic membrane by incident oscillatory variations in the air pressure originated in the vibrating body of the source of the sound [3]. As in the visual system, auditory sensations are characterised by different quantities that reflect fundamental properties of its physical sources. Loudness, like brightness, correlates with the amplitude of the oscillations [4]; pitch, like colour, correlates with its frequency [4].

Understanding the physiological mechanisms underlying perception is fundamental to comprehend how humans form an image of the world. Physiological studies investigating the peripheral organs responsible for gathering sensory information such as the eye, the ear, or the cochlea, revealed that key aspects of the sensory input are computed in the periphery; for instance, a large number of retinal cells are tuned to respond preferably to different electromagnetic wavelengths, effectively decoding colours from the incident light [3].

Similarly, the average firing rate of the hair cells in the cochlea is directly correlated with perceived loudness [5]. However, as we will discuss during this thesis, the computation performed by peripheral organs is not sufficient in order to explain how pitch is extracted from the oscillations of the tympanic membrane, indicating that the brain plays a fundamental role in pitch processing [5].

Pitch perception is indeed one of the fundamental open problems in auditory neuroscience [6]. Understanding pitch is essential to explain higher cognitive phenomena such as music or speech perception. Moreover, several auditory disorders like tinnitus [5] or a large part of the auditory processing disorder spectrum [7] arise from brain processing dysfunctionalities that cannot be studied without a comprehensive understanding of how the brain processes pitch.

This thesis addresses the study of the neural mechanisms underlying pitch perception, responsible for mapping oscillatory modes in the tympanic membrane into tonal sensations. We will argue that pitch is not single-handedly decoded in a specific part of the brain, but that pitch processing is a rather distributed computation implemented across several stages along the auditory pathway. Specifically, we will show that theories suggesting that pitch is integrally decoded in subcortical areas are incomplete, and we will introduce a novel theory

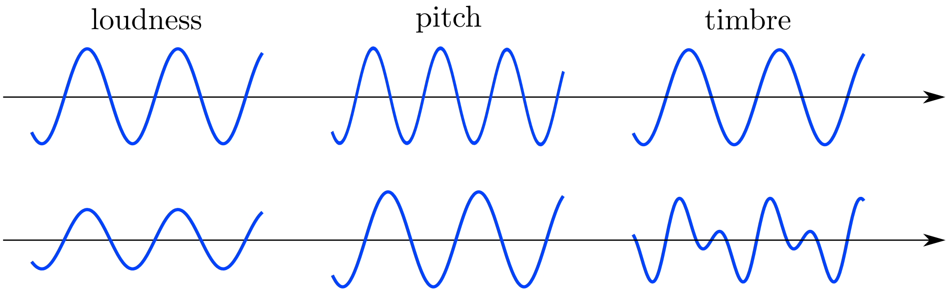


Figure 1.1: Representative sound waveforms illustrating the differences between loudness, pitch, and timbre. In each column, only the target dimension of the stimuli has been changed; for instance, the two waveforms in the third column show identical amplitude and repetition period and thus they elicit the same pitch and loudness perception but a different timbre.

of cortical pitch processing filling important gaps in the existing literature.

Our theory presents a comprehensive explanation of how pitch is processed in the auditory system and accounts, in a quantitative fashion, for electrophysiological observations that have intrigued auditory neuroscientists for several decades. Moreover, we will argue that the introduced pitch processing mechanisms can single-handedly provide for a parsimonious understanding of the origin of the higher order sensations of consonance and dissonance elicited by dyads comprising two simultaneous pitch values.

Contextual framework

Psychoacoustics

In psychoacoustics, the perception of tones is often considered as the result of combing three orthogonal components of the auditory sensation: loudness, pitch, and timbre [5]. In short, loudness is the perceptual correlate of the sound's waveform mean square amplitude, although it also depends on duration when the sounds are shorter than 200 ms [8]. Similarly, pitch is the perceptual correlate of the period of the fundamental oscillatory modes of the sound, if any: sounds whose waveforms present no periodicities at all do not elicit a pitch sensation [5]. The third dimension, timbre, is the perceptual correlate of the waveform's shape within a repetition period (see Figure 1.1). Timbre plays a crucial role in speech perception (for instance, vowels are uniquely characterise by their timbre) and in the identification of musical instruments [9].

In this work, we will use the term *single tone*[1] to describe any sound that can be uniquely characterised by its loudness, timbre and pitch. Under this definition, a dyad is a complex stimulus consisting of two simultaneous single tones and a melody is a succession of two or more single tones.

One of the main challenges in pitch modelling is that timbre and loudness are, up to certain extent, independent of pitch: very dissimilar waveforms can elicit the exact same pitch sensation [5]. In other words, the mechanisms underlying pitch perception should be blind to variations of loudness and timbre. Early studies in pitch perception carried out by Ohm and Helmholtz identified pitch with peaks in the Fourier spectrum of the sound's waveform [4] in a beautiful theory describing pitch as a linear phenomenon. This theory was

[1] This definition is chosen instead of the common *simple tone* to avoid a potential source of confusion with the term *pure tone*, used by some authors as a synonym of the former.

a natural evolution of the attributively Pythagorean model associating the pitch elicited by a vibrating string with the length of the string [10].

Physiological studies at the time discovered that different locations of basilar membrane, a structure in the cochlea that transforms mechanical vibrational modes into neural impulses, respond exclusively to specific oscillatory frequencies performing a sort of mechanical Fourier transform of the incoming sound waveforms [10]. However, further developments in psychoacoustics showed that pitch cannot simply explained via linear relationships between the oscillatory modes present in the stimulus. For instance, vibrating strings like the Pythagorean monochord are known to display several simultaneous oscillatory modes (eliciting several peaks in the Fourier spectra) but generally elicit a single pitch sensation [5]. Even more challenging, synthetic stimuli, such as the iterated rippled noises, elicit a very clear pitch sensation while presenting uniform Fourier spectra [11, 12].

The auditory pathway

Before pitch is processed in the brain, a first analysis of the incoming vibrational waves is performed by the peripheral auditory system. Peripheral processing can be regarded as a series of transformations mapping the sound's waveform into neural activity that propagates to the brain through the auditory nerve [5]. Neural activity at the auditory nerve is *phase-locked* to the sound's waveform, preserving the periodicities of the original waveform [10, 13].

The *auditory pathway* consists of all the neural areas directly or indirectly connected to the auditory nerve, comprising regions located in several hierarchical levels of subcortical and cortical regions [14]. Neural complexes along the auditory pathway are thus responsible for mapping the incoming phase-locked neural activity into a neural representation encapsulating our perception of pitch.

Subcortical structures present temporal properties that allow the auditory system to faithfully transmit phase-locked activity [15], whilst cortical regions are characterised by longer time constants. As a result, electrophysiological recordings reveal that phase-locked activity over ∼200 Hz is not observed in auditory cortex (AC) [16], indicating that periodicities in the neural activity of the auditory nerve are analysed along the subcortical pathway. This hypothesis is further supported by several studies identifying different candidate structures along the subcortical pathway that show selective activation to specific frequencies in the presence of periodic neural patterns [17, 18].

Thus, there seems to be mounting evidence that the analysis of periodicity, the physical property characterising pitch, is processed subcortically. However, functional data in human AC suggest that cortical regions play an active role in pitch processing: functional MRI studies found that only neural activity in AC correlates with the strength of the perceived pitch, suggesting that pitch salience is processed cortically [19]. Accordingly, several modelling studies conclude that AC performs some sort of integration of the subcortical processing, and that the longer time constants characterising cortical processing are crucial to derive a distinct pitch percept from subcortical representations [20, 21]. However, the specific neural basis of these processes are still unclear.

Cortical pitch processing

Magnetoencephalography (MEG), is a mesoscopic technique that measures the magnetic field variations elicited by collective post-synaptic potentials in cortical regions with a high temporal resolution. MEG recordings in human auditory cortex during pitch processing suggest a linear relationship between processing time and perceived pitch; specifically, time processing scales with $4T$, where T is the characteristic period of the elicited pitch [22, 23]. Periodicity detectors in subcortical areas require a whole repetition cycle in order to detect a given periodicity in the stimulus implying a linear relationship between processing time and pitch value, but suggesting a factor 1 rather than 4 in such dependency [17]; such divergence in processing time seems to indicate that cortex plays an active role in pitch processing.

MEG data successfully captures collective cortical dynamics with a relatively coarse spatial resolution, ignoring possible lower-scale phenomena that might be important to describe pitch processing [24]. Recordings using local field potentials (LFP), an intracranial version of electroencephalography (EEG), display a greater signal-to-noise ratio and a higher spatial resolution than MEG [25]. LFP studies reveal that several subdivisions of AC collaborate with each other during pitch processing, implying that cortical pitch processing is carried out by a distributed network of cortical areas within Heschl's gyrus and planum temporale [26].

fMRI, E/MEG and LFPs describe neural activity in a mesoscopic scale. Higher resolution methods exploring microscopic neural dynamics require invasive interventions and are extremely rare in humans. However, several studies have found a certain degree of functional similarities between human AC and the cortical counterparts in other mammals [14], suggesting that some results from animal research are general enough to be applied in the human domain. intracranial data in Marmosets [27] indicates that distinct pitch values are encoded in the activity of groups of at least 10 neurons in cortex, implying that collective rather than individual neural behaviour is responsible for pitch coding. This result suggests that the appropriate scale to characterise cortical pitch processing is the mesoscopic scale, supporting the use of MEG data in our investigation.

Research objectives addressed in this thesis

The main objective of this thesis is to develop a theoretical model describing the neural mechanisms responsible for cortical pitch processing in human auditory cortex. This model should be able to explain multiple neuroscientific questions regarding the nature and functioning of auditory cortex, and account for the experimental results shown above. This general objective can be subdivided in several research objectives that are detailed in the following paragraphs.

1. Description of the neural representation of pitch along the multiple stages of the auditory pathway. A neural representation is the specific neural code mapping a piece of information in a *neural complex* [28]. For instance, at the very beginning of the auditory pathway, the spectral properties of the sound relevant for pitch extraction are encoded in the phase-locked activity of the auditory nerve elicited by the peripheral auditory system [5].

Hierarchical theories of neural organisation suggest that sensory systems are structured in a hierarchy of neural levels, each of them responsible for transforming the output of the previous level into a more abstract representations [29]. The auditory pathway is believed to follow a hierarchical structure of this kind, comprising multiple stages from the beginning of the auditory nerve up to auditory cortex [14]. Understanding the representation of pitch at each stage of this hierarchy is an important pre-requisite to reveal the mechanisms responsible for cortical pitch processing.

2. Elucidation of the neural mesoscopic mechanisms underlying cortical pitch processing. Following the hierarchical scheme outlined above, we can define cortical pitch processing as the transformation between the neural representation of pitch in the last subcortical relay in the auditory pathway and the more abstract cortical counterpart. This question addresses the problem of how neural complexes in cortex perform such transformations. Based on the hypothesis that pitch processing is a collective phenomenon, we will approach this problem from a mesoscopic rather than a microscopic scale, describing cortical regions as concentrations of populations of neurons and ignoring the dynamics of individual cells inside each neural ensemble [30].

3. Explanation of how cortical pitch processing gives rise to the observed dynamics in MEG recordings. MEG recordings show that regions in human auditory cortex exhibit specific and stereotypical dynamics during pitch processing. These evoked fields are elicited by the aggregated dynamics at the population level in cortex [24]. Our objective is to use our model to understand the origin of the elicited fields, and to use the experimental data in order to test and validate our model.

In order to solve this problem, we will first study how trends observed in the MEG data can be generated by the neural populations drawn in the cortical model of our theory. Afterwards, we will try to quantitatively explain relevant properties of the data .Specifically, our aim is to explain the observed correlation between the latency of MEG transient responses to pitch onset and perceived pitch.

4. Illustration of the role of the different anatomical regions in human auditory cortex. The auditory cortex is defined as the set of cortical regions that receive direct input from subcortical regions of the auditory pathway. Source localisation in MEG data [22] and LFP studies [26] further identify several parts of AC that selectively activate during pitch processing. Using our model, we aim to explain the functional structure of those pitch-related cortical areas and establish their hierarchical organisation; i.e, to study whether the specific cortical regions belong at the same hierarchical level and, in case a hierarchy exists, to discern whether the information is simply transmitted from the lower to the higher level in a bottom-up fashion or whether they also display top-down interactions [31].

5. Extension of the model to consider the processing of sounds eliciting multiple simultaneous pitch sensations. Single tones with different characteristic pitch values show complex interactions when processed simultaneously, giving rise to a new perceptual dimension often described as consonance and dissonance [4]. Our last research question addresses the origin of such interactions by combining novel experimental data in dyads with our theory of cortical pitch processing. Our aim is to investigate if pitch processing in dyads is processed in a linear way; i.e, if it can be described as two parallel, independent processes. If that is the case, consonance and dissonance perception might be the result of a later process carried out at a higher level of the auditory pathway. On the contrary, if singlets interact in a non-linear fashion during pitch processing, it would be interesting to test if the dissonance and consonance percepts are a consequence of such interaction.

Structure of the thesis

This thesis is structured as follows. Chapter 2 summarises our literature review in auditory experimental neuroscience providing for context, restrictions, and clues for the development of the cortical model.

Chapter 3 shows the state of the art in models of pitch processing both in subcortical and cortical areas. In this chapter, we will argue that auditory cortex plays an essential role in pitch processing, and that top-down interactions between different cortical areas are essential to explain perceptual and neuromagnetic results on the processing of complex stimuli. Moreover, we will show that previous models of pitch processing were not biophysically specific enough in order to understand the underlying mechanisms of pitch processing in auditory cortex.

Chapter 4 introduces a general model of pitch processing involving several cortical areas and describing the hierarchical structure of the pitch-related auditory pathway. The proposed mechanisms explain the dynamics of crucial components of the evoked fields.

Chapter 5 expands our previous results to the study of dyads and the processing of consonance and dissonance in human auditory cortex.

Finally, Chapter 6 comprises the concluding remarks, a recapitulation our results, final conclusions, and the explicit evaluation of the research questions detailed above.

Chapter 2

The puzzle of auditory processing

State of the art knowledge on auditory processing consists of a collection of psychoacoustic phenomena, brain-imaging and electrophysiological data, that complement each other revealing key aspects of the mechanisms underlying the phenomenon of pitch. However, the relationship between neural, anatomical and functional brain-imaging responses and psychoacoustics have been puzzling researchers for decades. In this chapter, we will review and contextualise the most relevant findings in order to establish an experimental framework for our cortical theory.

Anatomy of the auditory system

The auditory system anatomy is almost identically replicated in right and left hemispheres [3]. All the structures described in this section are thus implemented twice: mammals present two peripheral systems, two inferior colliculi, two auditory cortices, etc.; although lateral specialisation has been found in high-level cognitive processes such as music or speech perception in cortex [1, 32–35], structures in the subcortical pathway seem to be essentially symmetric.

Peripheral auditory system

Anatomy of the peripheral system

The peripheral auditory system transforms the mechanical oscillations carrying the sound frequencies into a neural representation. It consists of three main structures: the outer, the middle and the inner ear [5]. The most important element in the outer ear is the tympanic membrane, a thin cone-shaped membrane that oscillates in synchrony with the local pressure variations carrying the sound's waveform [5]. Oscillatory modes at the tympanic membrane are filtered and transmitted by a set of three ossicles located in the middle ear to the oval window in the cochlea (see Figure 2.1A) [5], that adapt the low impedance of air to the impedance of the inner fluid in the cochlea, approximately 4000 times higher [36].

Cochlear temporal processing and phase locking

The cochlea is located in the inner ear and presents a spiral-like coiled formation consisting in two fluid-filled chambers separated by the *basilar membrane* (BM, see Figure 2.1) [5].

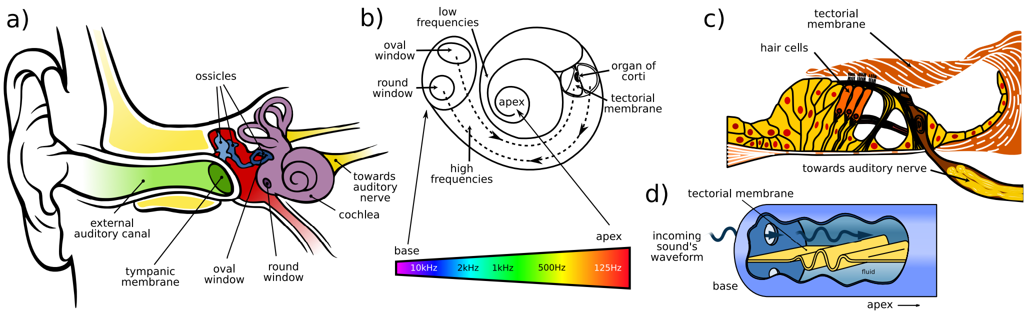


Figure 2.1: Schematics of the anatomy of the peripheral auditory system at different levels of detail. a) Outer, middle, and inner ear, (adapted from [37], Fig. 1); b) Structure of the cochlea (adapted from an illustration in the public domain); c) Detail of the Organ of Corti (adapted from an original illustration by commons.wikimedia.org/wiki/User:Madhero88); d) Detail of the sound propagation along the cochlea and the basilar membrane (adapted from an illustration in the public domain).

Auditory vibrations are transmitted from the oval window in the outermost part of the cochlea or base to the basilar membrane through the fluid-filled chambers [5]. The BM is in contact with the organ of Corti, populated by hair cells that activate in synchrony with the mechanical movement of the membrane [5]. Hair cells are endowed with a series of stereocillia whose displacement provokes a short neural pulse known as *spike* [3].

Stereocillia of different lengths and threshold populate the organ of Corti, ensuring a dynamic response to the deflections of the BM [3]. Generally, a strong burst of activity is observed during the instants of maximum depression of the membrane, whilst low levels of activation are displayed during the remaining of the oscillatory cycle (see Figure 2.2). Neural activity at this point is thus *phase locked* to the stimulus' waveform, preserving all the spectral information necessary to decode pitch [5]. Moreover, since louder sounds provoke larger displacements of the BM, hair cells firing rates also reflect the intensity of the stimulus [5, 38].

Robust phase-locked activity in the mammal auditory nerve has been experimentally measured up to 3 kHz–6 kHz, depending on the species [13]. Over the phase-locked limit, neurons are unable to recover from the previous spike before the next phase occurs, which results in a gradual degradation of transmission fidelity [5]. The actual phase-locked limit in humans can only be measured with invasive techniques and is still unknown [40]; although some studies argue that it might be as high as 5000 Hz [41], more conservative estimations establish the limit at around 2000 Hz [42].

Cochlear spectral processing and tonotopy

As shown in Figure 2.2, not all the hair cells along the basilar membrane respond equally to a given stimulus. In fact, as consequence of a stiffness gradient presented along the basilar membrane, different locations along the membrane respond preferably to certain frequencies [43]. The edge of the BM closest to the termination of the ossicles or *base* responds preferably to high frequencies; whilst the innermost edge, located at the centre of the coil and termed *apex*, is tuned to favour low frequencies [43] (see Figure 2.1). Nerves coming from hair cells located at different positions along the basilar membrane are typically referred to as *cochlear channels*. Since the propagation speed within the cochlea is finite, cochlear channels at the base, encoding high frequencies, respond up to 11 ms earlier than the innermost cochlear channels at the apex [5].

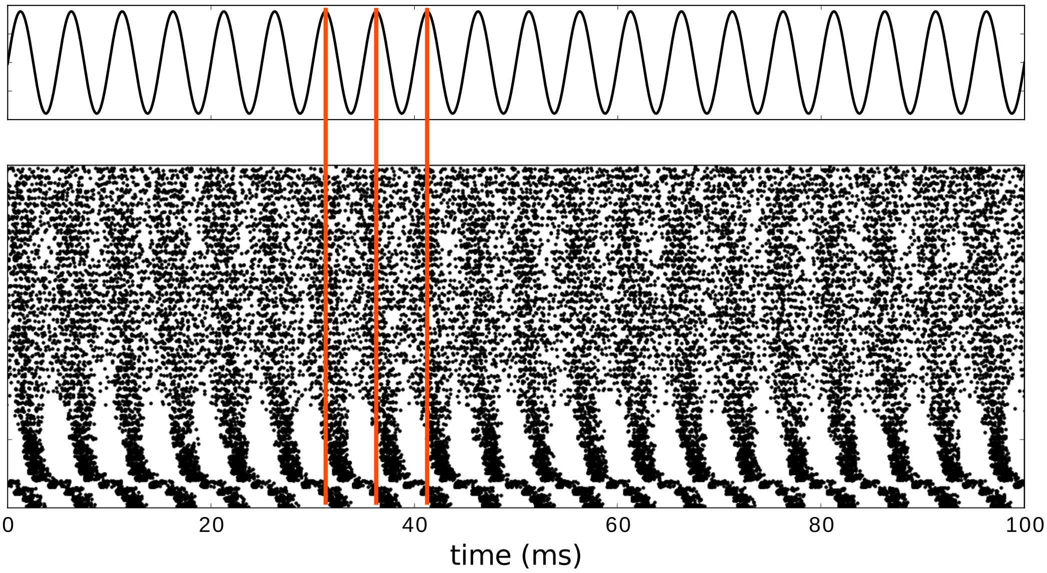


Figure 2.2: Raster plot of the neural cochlear spike trains generated by a pure tone. Top panel shows the stimulus' waveform, a pure sinusoid with a 200 Hz frequency; bottom panel shows the spikes elicited in different hair cells along the basilar membrane by the stimulus. Spikes trains were simulated using a recent model of the auditory periphery [39].

Neural activity in the early auditory nerve consist of a set of phase-locked spike trains transmitted through different cochlear channels [5]. Spectral information at this stage is said to be represented in two coexisting neural codes: a *temporal code*, consisting of the fine-grain temporal information of the independent spike trains, and a *place code*, in the mean activity of the different cochlear channels. Place code is found in several stages along the auditory pathway; this arrangement is known as *tonotopy*, meaning tonal topology.

Although not explicitly noticeable in the raster plot in Figure 2.2 due to the overlapping of spikes during the phase-locked bursts of activity, both, place and time codes, could be indistinctly used to extract pitch of pure tones. However, experiments with more complex stimuli show that tonotopic information by itself is not enough to infer the elicited pitch (e.g. [44]), whilst phase-locked information preserves pitch-related information in a more robust, timbre-independent, fashion (see Figure 2.3).

Subcortical pathway

The subcortical auditory pathway is a complex comprising a vast number of bodies and substructures receiving direct or indirect input from the cochlear hair cells [14].

Essential bodies of the subcortical pathway

Microscopic functional imaging of the brain requires invasive techniques, and the anatomical structures along the auditory subcortical pathway display a relatively large similarity across mammals [14]. Most of the results listed below derive from animal studies.

The subcortical auditory pathway consists of a series of relays organised in a hierarchical way. In ascending order, the most significant bodies for our investigation are (see Figure 2.4A): 1) the cochlear nucleus complex (CN), 2) the superior olivary nucleus (SOC),

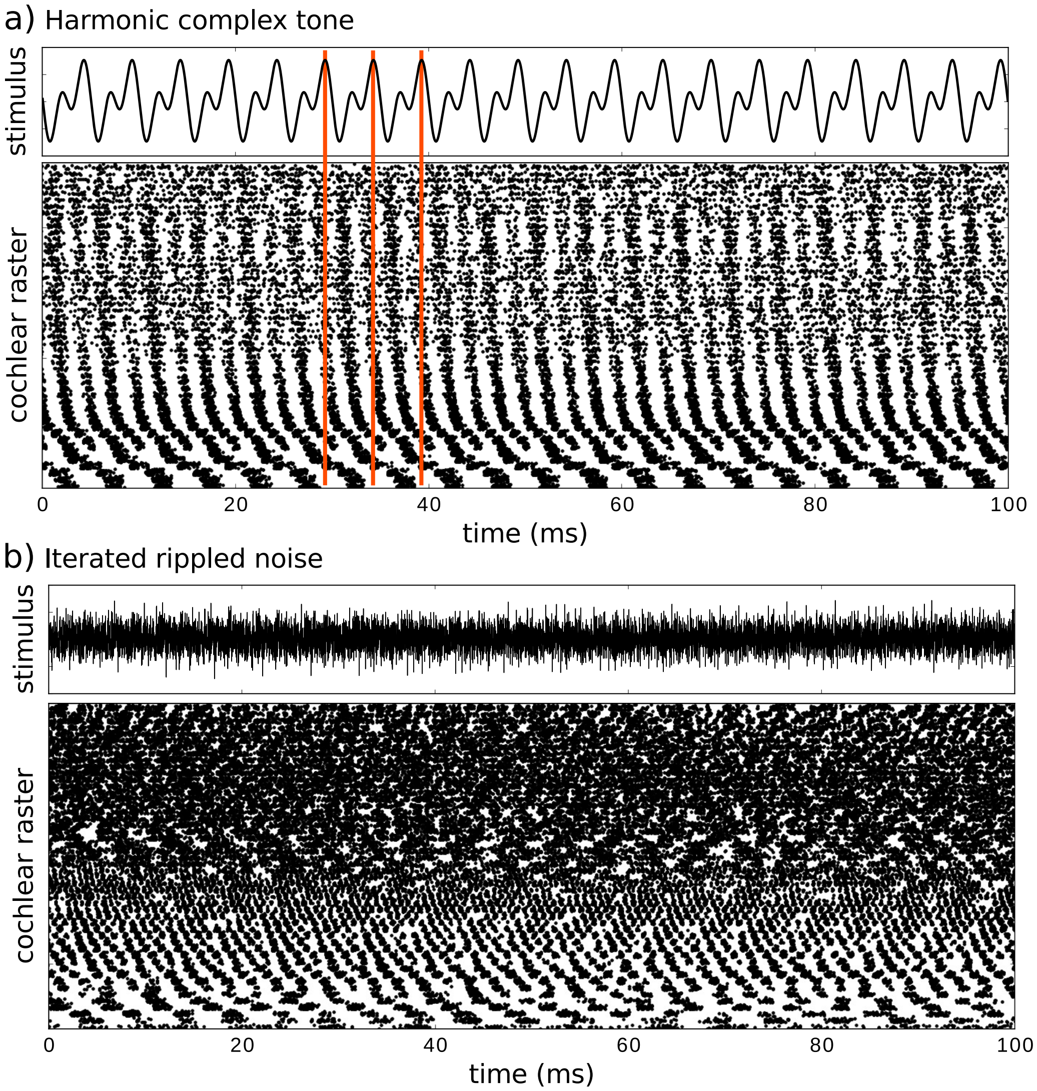


Figure 2.3: Raster plot of the neural cochlear spike trains generated by a harmonic complex tone and an iterated rippled noise. a) Stimulus was a harmonic complex tone with two harmonics and a fundamental frequency of 200 Hz. b) Stimulus was an iterated rippled noise with a delay of 5 ms and 16 iterations; cochlear responses were simulated using a model of the auditory periphery [39].

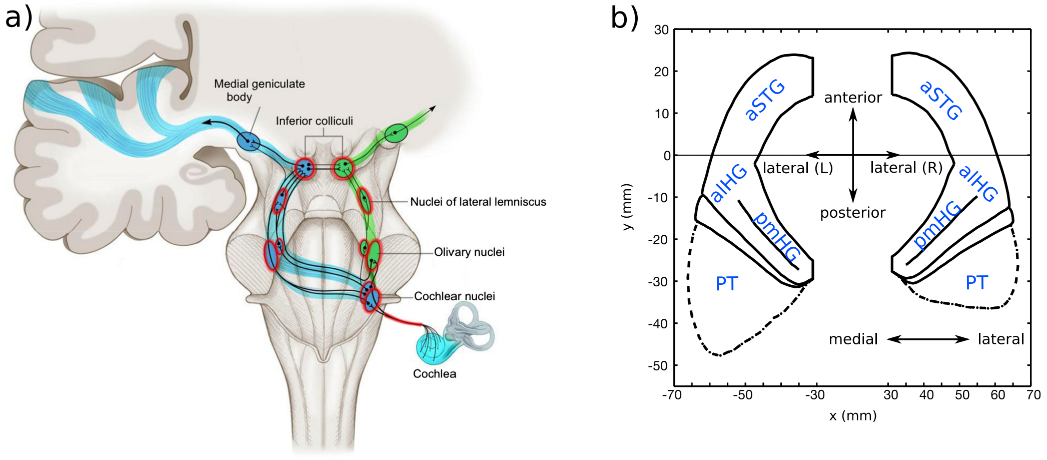


Figure 2.4: Anatomical schematics of the auditory brain. a) The ascending auditory pathway. Figure adapted from the *Ear Anatomy* series by Robert Jackler and Christine Gralapp. b) Schematic view of the human auditory cortex in Tailarach coordinates. Schema adapted from [45], Fig. 9.

3) the inferior colliculus (IC), and 4) the medial geniculate body (MGB) [14] . The specific functional role of each of these complexes is not fully understood.

Neural connectivity along the subcortical pathway

Hair cell activity is first transmitted to the cochlear nucleus. The CN is connected to the olivary nuclei, which connects with the inferior colliculus; the later connects with the medial geniculate body, that outputs to auditory cortex [14]. Besides these *bottom-up* connections, a descending pathway also connects these bodies in a *top-down* manner. Bottom-up processes transmit the auditory cochlear inputs to higher stages of the auditory pathway holding more abstract representations [14]; top-down processes operate in the reverse way, modulating way in which the lower processing centres transform the information [5]. For instance, top-down afferents are known to modulate the amplification of the hair cells in the organ of Corti [14].

The arrangement of the auditory nerve along the subcortical pathway preserves cochlear tonotopy [14, 46] (see Figure 2.5); moreover, primary subcortical cells present temporal properties ensuring a high fidelity transmission of the phase-locked incoming spike trains [15].

Evidence of subcortical spectral processing in mammals

The IC is mainly populated by a type of disc-shaped principal cells arranged in such a way that their terminations form a series of sheets presenting a double spectral organisation [47]: one dimension spans the tonotopic arrangement as discussed before; a second intra-laminae structure follows an isofrequency contour with a *periodotopic* shape; i.e, a period-based topology [15], reflecting spectral properties of the phase-locked spike trains generated at the cochlea [47]. Lateral inhibition across the first, tonotopic dimension, sharpens the responses to specific frequencies [15].

Although this tonotopic-periodotopic laminar organisation seems to be common to all mammals, the distribution of frequencies is different across species [47], suggesting that the organisation of IC reflects the spectral resolution of the auditory system. For instance, one third of the IC of the moustache bat is specialised in frequencies around 60 kHz, which are essential in echo-location [47].

Cells at the medial geniculate body receive ascending input from both, the inferior colliculus and the cochlear nucleus, and outputs to the primary auditory cortex [48]. There is evidence for two different tonotopic fields in the MGB, the first field spans a large frequency range from 7 Hz up to 16 kHz; the second presents sharper tuning curves and seems to reflect the tonotopic organisation of the IC [48].

Evidence of subcortical spectral processing in humans

Functional magnetic resonance imaging (fMRI) is a popular, non-invasive brain imaging technique, that measures the amount of oxygenated blood in brain tissues in order to detect functional centres of activation in time [49]. fMRI techniques present a very poor temporal resolution, with refreshing times of the order of a few seconds, but a reasonably good spatial 3-dimensional resolution, of the order of the cubic millimetre [49], that can be exploited to find mesoscopic areas active during pitch processing.

fMRI data recorded during pitch processing in human subjects shows that overall activation at the cochlear nucleus and the inferior colliculus correlates with the degree of regularity of the sound, presenting higher levels of activation for sounds eliciting stronger pitch sensations [50]. However, these results were not replicated in later study [19]. None of the studies reported a significant correlation between activation of the MGB and pitch strength [19, 50].

Tonotopicity in the human MGB was confirmed by recent high-resolution fMRI data [51]. Although highly likely, the tonotopic arrangement of the human IC and CN has still not been reported due to the small size of these structures relative to the resolution of the current fMRI technology.

Auditory cortex

Auditory cortex (AC) is defined as the area of the temporal lobe receiving inputs from the medial geniculate body [52]. The human AC is subdivided in primary and secondary auditory cortices. Primary auditory cortex (A1) occupies the Heschl's gyrus (HG) [16] and extends up to regions of planum temporale (PT) [53]. Secondary auditory cortex (A2) is located next to A1 and also comprises adjacent areas of planum polare and planum temporale [16] (see Figure 2.4B).

Auditory cortex outputs to higher cortical stages located in frontal and temporal lobes [52], related with higher order cognitive functions and cross-sensory integration [54].

Structure of the auditory cortex in humans and mammals

Tripartite organisation of auditory cortex. Mammals show a tripartite auditory cortex consisting of a *core*, a *belt*, and a *parabelt*; these substructures putatively show an increasingly abstract representation of the auditory stimuli [52]. The core presents a tonotopic arrangement [52, 55] and a strong phase-locked activity to low-frequency stimuli up to 300 Hz [55–58]. Cortical neurons responding preferably to specific stimulus' frequencies are predominantly common in the core-belt complex in marmosets [15, 56].

The human AC presents a similar organisation. The Heschl's gyrus, enclosing most of A1, is divided in its posteromedial (pmHG) and anterolateral (alHG) sections, which present similar properties as the core and the belt, respectively [15, 16]. Similarly, A2 is related to the mammal parabelt [52]. An intracranial study in humans [16] showed that whilst posteriomedial Heschl's gyrus presents reliable phase-locked activity up to 50 Hz, its fidelity gradually faints over 50 Hz and completely vanishes over 200 Hz. This pattern parallels the behaviour of marmoset's cortical neurons that present a faithful phase-locked response up to 100 Hz and a gradual decay up to the limit of 300 Hz [56]. The human alHG does not show phase-locked activity at any frequency range [16].

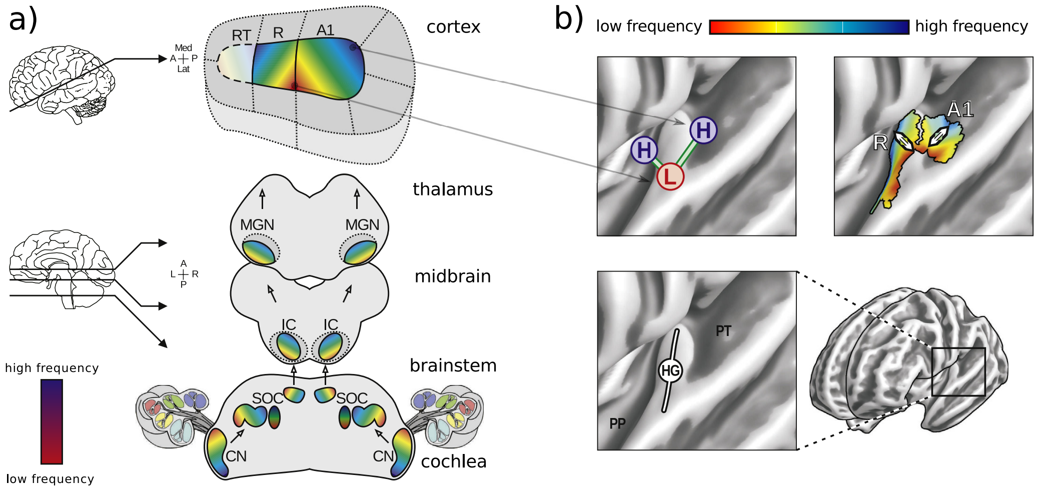


Figure 2.5: Tonotopy along the auditory system. a) Tonotopic arrangement of the ascending auditory pathway. b) Anatomic detail of the two adjacent tonotopic maps found in auditory cortex. Figures were adapted from [46], Fig. 1 and 2.

Functional connectivity within A1. Adjacent areas of HG show strong *bottom-up* and *top-down* nerve afferents that seem to indicate a close collaboration between nearby areas during stimulus processing [52]. Accordingly, top-down and bottom-up functional connections have been found between pmHG and alHG during pitch processing in intracranial recordings of local field potentials (LFP) [26].

Secondary auditory cortex and higher order processing. fMRI studies suggest that A2 is specialised in high-order cognitive tasks such as music or speech processing [50]. Unlike previous auditory structures, A2 shows lateralisation effects: most subjects show a larger activation in the right A2 during music processing and a larger activation in the left A2 during speech processing [32, 33, 50].

Top-down connections between primary and secondary auditory areas have also been reported in fMRI studies, that show A2 engagement during low-level cognitive processes carried out in A1 [19].

Spectral representations in auditory cortex

Tonotopy in cortex. Although intracranial recordings in humans failed to find a tonotopic neural arrangement in human AC [16], larger scale fMRI studies have systematically found two contiguous tonotopic maps in HG and the adjacent rostral field [46, 54] (see Figure 2.5). Some sections of the tonotopic map in A1 show lateral inhibition effectively sharpening the frequency contours of the maps in cortex [59].

Periodotopy in cortex. Cortical periodotopic arrangements as observed in IC have only been found in cats' AC core [57]. However, since phase-locked activity in auditory cortex vanishes over 200–300 Hz [16, 56–58], temporal information must be encoded in some sort of rate code like the cat's periodotopic field in other mammals in order to explain to explain how complex stimuli elicit a pitch percept that cannot be accounted for by the tonotopic code alone [60].

Moreover, tonotopy and periodotopy might have been confounded in traditional fMRI setups that use as stimuli pure sinusoids eliciting identical tonotopic and periodotopic maps, both of them matching the evoked pitch [46]. fMRI studies using stimuli with more complex spectra found high correlations between periodotopic maps elicited by natural sounds and tonotopic maps elicited by sinusoids in cortex [54].

Harmonic representations High resolution fMRI studies have found that selective parts of the regions responding preferably to a given frequency f_0 co-activate as well when presenting stimuli with harmonically related fundamental frequencies (e.g. $2f_0$ or $f_0/2$) [59,61]. Harmonic co-activation of frequency-tuned neurons has been largely reported as well in other mammals [62].

Are both, place- and time-codes, necessary for cortical pitch processing? Frequency tuning of neurons in the human auditory cortex shows a much higher resolution, up to an eighth of an octave, than the spectral resolution of the basilar membrane [63]; this suggests that the time-code rather than the place-code is responsible for cortical spectral analysis. However, intracranial studies in marmosets [64] and psychoacoustic experiments in humans [41,60] indicate that neural activation correlates with resolvability properties of the place-code. Thus, evidence seems to imply that both representations play an important role in pitch processing.

Evidence for cortical pitch processing

Pitch-selective activation is found in early areas of the subcortical auditory pathway [50], suggesting that pitch is processed between the cochlear nucleus and the inferior colliculus. However, subcortical activation in fMRI studies does not vary when presenting sounds with different pitch strengths [19]. Correlation with pitch strength is however found in alHG [19], indicating that pitch is partially processed in AC.

Pitch-selective neural populations have been consistently reported in a region overlapping the low-frequency tonotopic region of the human alHG [19,50,54,65]. This aggregate of pitch-selective neurons has been labelled as the cortical *pitch centre* [57], a neural ensemble hypothetically responsible for cortical pitch processing. Intra-cranial studies in awake marmosets also found a localised set of pitch-selective neurons in regions analogous to alHG using stimuli with different spectral envelopes [27,66], supporting the idea of a pitch processing centre in the mammal auditory cortex. Further results in MEG and EEG studies also converge on the idea of a putative pitch centre in alHG–PT (see §2.3).

On the contrary, some studies argue that cortical pitch processing might be timbre-dependent, suggesting that different centres across the AC could be responsible for pitch processing, depending on the sound's spectral shape [15]. Accordingly, a study in the ferret auditory cortex reported that pitch-selective neurons were scattered along five different areas of AC and that they did not only reflect pitch preference, but also responded selectively to timbre and localisation of sounds [67]. However, the identification of pitch-selective neurons in animal intracranial data is a highly challenging task; for instance, neurons responding to the animal choice derived from the evoked sensation could be indistinguishable form neurons holding or processing the pitch itself [68]. More generally, pitch-responsive neurons are not necessarily involved in pitch processing.

Another common argument against the idea of a generalised processing carried out by a single putative pitch centre in human AC is the divergence on its exact location across different studies, ranging from alHG to adjacent areas of planum temporale [53]. Cortical tonotopic maps themselves show a large subject-to-subject variability, indicating that locations within AC are difficult to compare across subjects [46]. Nevertheless, a key study in fMRI [65] shows that measuring the tonotopic map in a subject-like fashion leads to a consistent anatomical location of the pitch-selective cortical centre, indicating that the variability of the results might rely on inter-subject, rather than inter-stimuli, variations.

Psychoacoustics

Psychoacoustics is the branch of psychology studying the phenomenology of hearing [5]. In this section, we will introduce some basic results on the psychoacoustics of single tones, defined as sounds that can be uniquely characterised by three approximately independent properties: loudness, timbre, and pitch.

Loudness

Loudness is the perceptual correlate of sounds' intensity and, for stimuli longer than 200 ms [8], it scales linearly with the logarithm of the square amplitude of the sound's waveform [38]. As such, loudness is most likely encoded in the time-integrated overall firing rate of the hair cells at the organ of Corti [5].

Loudness is commonly measured in *decibels (sound pressure level)*, or dB SPL, defined as $s = 10 \log_{10} I/I_0$ where I_0 is the average threshold under which sounds are no longer audible. Intensity levels over 90 dB SPL can lead to long-term hearing loses; 140 dB SPL is consider as the threshold of pain [5].

Timbre

Timbre is the perceptual correlate of the shape of the sounds' waveform and it encompasses temporal aspects of the stimulus not related to its fundamental period [9]. Rescaling a given spectral shape towards higher or lower frequencies produces a change in pitch, but not in timbre. Timbre is used to characterise the sound of musical instruments [9]; vowels are also characterised by their elicited timbre [9].

Timbre can be resolved in sounds as short as 2 ms [69], indicating a much faster processing than pitch [70] (that usually take around 4 times the period of the waveform, e.g. $\sim$10 ms for a 250 Hz tone).

Pitch

Pitch is the perceptual correlate of the period of the stimulus' waveform; i.e, the fundamental frequency f_0 of the present oscillatory modes [43]. Waveforms showing no periodic elements do not evoke a pitch sensation; waveforms with less defined periodic modes elicit weaker pitch sensations [43].

Harmonic complexes

The simplest periodic waveform is the sinusoid or *pure tone* (PT). A pure tone evokes a pitch corresponding to its frequency of oscillation [4]. Pure tones are synthesised in laboratories and are often used in hearing research, but they are not common in nature. Rather, natural sounds present a more complex spectral shape, often consisting on several overlapping sinusoids.

When all the overlapping sinusoids in a complex are harmonically related, i.e, the frequencies of all the overlapping oscillation modes are multiple of a common ground frequency f_0, the complex is called a *harmonic complex tone* (HCT) [5]. Vibrating strings like the monochord's generate waveforms belonging to this category [4]. HCTs elicit the same pitch sensation as a sinusoid with frequency f_0, although attending experienced listeners have reported to be able to perceive several simultaneous pitch values associated to the different harmonics comprised in the complex [4].

The strength of the pitch sensation evoked by HCTs increases with the number of harmonics present in the complex [71], and decreases dramatically when the higher order harmonics span high-frequency ranges where the basilar membrane presents a lower spectral resolution, and are thus not independently resolved by separated cochlear channels [72].

A *click train* is a limit case of an HCT when all harmonics of the harmonic series are present in the stimulus [73], also eliciting a strong pitch sensation equivalent to f_0 . Click trains can equivalently be described as a succession of Dirac's deltas or *clicks* separated by a constant $t = 1/f_0$. Random perturbations in the inter-click intervals can be used to decrease the strength of the evoked pitch [74].

Virtual pitch

Harmonic complex tones evoke the pitch percept of its fundamental frequency f_0 even if the fundamental itself does not form part of the complex [60]; this phenomenon is known as *virtual pitch*, and has been observed in other mammals such as monkeys, marmosets [64] or cats [57].

Although virtual pitch has been suggested to be a side effect of non-linear interactions at the basilar membrane eliciting activation at cochlear channels characterising f_0, this hypothesis was disproved by introducing a masking element shunting elicited activation around those cochlear channels [5]. Moreover, a short sequential display of the independent harmonics, in such a way that they are separately present in the cochlea, also elicits a virtual pitch sensation [75].

Waveforms of HCTs with a missing fundamental are still periodic in f_0, supporting the hypothesis that pitch is decoded using a time-code. To test this theory, a study [60] used alternated-phase HCTs (ALT HCTs), where the phase of the odd-numbered harmonics is reversed with respect to the phase of the even-numbered ones. ALT HCTs show identical long-term Fourier spectra as their phase-aligned counterparts, but their waveform is periodic in $2\,f_0$ rather than in f_0 [60]. Results of the study showed that, when harmonics in the ALT HCT are resolved in separated cochlear channels, the stimulus elicits a pitch equivalent to f_0; when the harmonics are not separately solved in the basilar membrane, the complex evokes a $2\,f_0$ pitch [60]. Consistent effects have been observed in marmosets [64].

Iterated rippled noises

Iterated rippled noises (IRN) [11,12] are a kind of auditory stimulus consisting of the aggregation of a series of repetitions of a given sampled white noise iteratively delayed by a fixed period δt. Although IRNs present flat long-term Fourier spectra, they elicit a pitch sensation equivalent to $f_0 = 1/\delta t$ [11]; the strength of the sensation scales up with the number of recursively delayed repetitions [12]. IRNs can elicit a pitch even if they consist of only two iterations delivered separately in different ears [43], and elicit a varying pitch percept when the delay δt is varied smoothly [76].

Thresholds

Periodic sounds with fundamental frequencies between 30 Hz [77] and 4–5 kHz [41] elicit a sensation of pitch in humans, although pitch discrimination seems largely reduced above 3–4 kHz [43]. Pitch sensations can be elicited by tones as short as two of their oscillatory periods, but durations of over four cycles are necessary for robust pitch discrimination [43].

Subject-specific phenomena

Pitch perception presents a large inter-subject variability. Discrimination thresholds follow similar shapes in different subjects, but the magnitudes can vary widely depending in genetic factors, age, and experience [10].

Absolute pitch. Subjects with absolute pitch are able to label the pitch of a sound without using a reference tone [78]. In contrast, most of the listeners need to match target sounds with a sample tone to assign it a pitch label; i.e, they perceive pitch as a relative, rather than absolute, property [5]. Evidence suggests that absolute pitch cannot be learned nor trained [5, 78].

Spectral and fundamental listeners. Another subject-specific is observed when considering some intervals of HCTs with a missing fundamental. If the lowest harmonic of one of the complexes presents a higher frequency than the lowest harmonic of the second tone, but the fundamental frequency of the first tone is lower than the fundamental of the second (see Figure 2.6A) [4]. Some subjects, called *spectral listeners* (f_{SP}), are more likely to judge the pitch difference between the two tone according to the frequency of the lowest present harmonics, whilst other subjects, called *fundamental listeners* (f_0), often judge the pitch change accordingly to the f_0 of the HCTs (see Figure 2.6B) [4, 5]. Intriguingly, f_{SP} / f_0 preference is correlated with the relative size of the lateral portion of Heschl's Gyrus, suggesting a hemispherial lateralisation effect in pitch interval judgements [45].

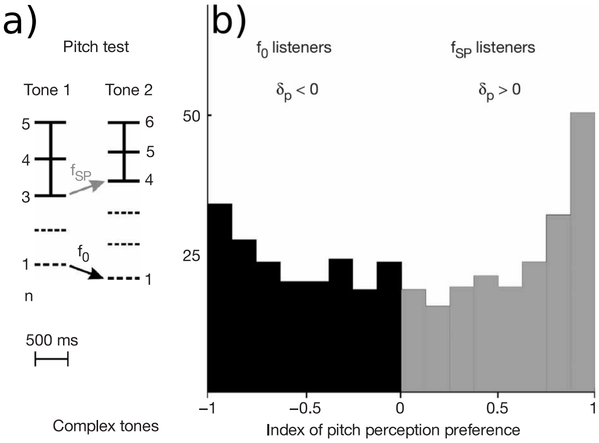


Figure 2.6: Fundamental and spectral listeners. a) Schematic representation of the stimuli. b) Distribution of subjects (of a sample of 420) across the continuous line between pure-f_0 and pure-f_{SP} listeners. Figure taken from [45], Fig 1.

Electrophysiology

Most neural communication is conveyed by means of short electrical impulses, called *action potentials* or *spikes*, that propagate along the source cell axon towards the target neuron [3]. Aggregations of nearby spiking neurons with similarly oriented axons generate a net electric current that is strong enough to have a noticeable effect on the electromagnetic fields at the scalp [79]. Electroencephalography (EEG) and magnetoencephalography (MEG) measure such elicited fields in order to infer mesoscopic neural activity in the brain in a non invasive fashion. Due to its exquisite temporal resolution, E/MEG imaging plays a crucial role in auditory neuroscientific research.

The physiological basis of E/MEG

Equivalent dipoles

Action potentials propagate along an axon following a succession of ion exchanges between the neural body and its extracellular fluid [80]. Ionic currents elicit local electric and magnetic fields [80] that propagate through the cortical tissues all the way to the scalp. Although field variations provoked by a single neuron dissipate on the large ocean of brain electric activity, aggregated fields elicited by the collective activation of a neural ensemble are strong enough to display relatively large signal to noise ratios in the scalp [80].

When the distance between the axons carrying the electric current is much smaller than the distance from the axons to the electromagnetic sensors, the aggregated source of the fields can be approximated by an equivalent magnetic dipole **d**, represented as a vector with the direction of the electric flow (from source to sink) and the magnitude of the electric current [81]. Equivalent dipoles are good approximations of the fields elicited by a neural population when the source-to-sink distance is finite and smaller than a few millimetres [82].

The human cerebral cortex is a thin, folded, tissue, anatomically organised in six consecutive layers, I (near the scalp) to VI (closer to thalamus). Cortical regions are often studied as vertical columns encompassing these six layers [3]. Intra-neural currents flow in parallel to those columns, eliciting an equivalent dipole orthogonal to the cortical surface [82].

A fold or ridge in the folded cortical tissue is called a *gyrus*, and a groove a *sulcus* [3] (see Figure 2.7a). Auditory cortex, located in Heschl's gyrus, elicits both, radial and tangential dipoles [82] (see Figure 2.7b). EEG is sensitive to both of them, but MEG is blind to radial dipoles and it only measures tangential sources [24].

Nearby dipoles (i.e, dipoles much closer to each other than to the sensor) aggregate into equivalent dipoles summarising the overall elicit field according to the superposition principle [81] (see Figure 2.7c).

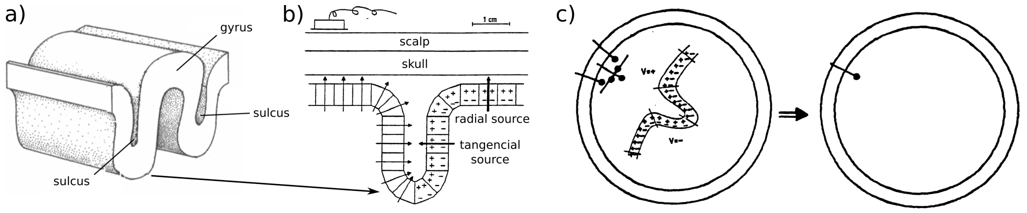


Figure 2.7: Equivalent dipole orientations in Cortex. a) Schematic view of gyri and sulci in the folded cortical tissue (Figure from the public domain). b) Dipole orientations across a cortical sulcus (Figure adapted from [82], Fig. 1). c) Example of how nearby dipoles aggregate into an equivalent single source (Figure adapted from [82], Fig. 6).

Dipole fitting

Given a certain magnetic dipole result of the spiking activity of a neural population, the electromagnetic fields elicited in the scalp surface are calculated using the Poisson equation and an anatomic model of the head [82]. The solution to this problem is unique and its precision is only bounded by the exactitude of the head model [24].

On the contrary, the *inverse problem* of inferring the location and strength of an equivalent dipole given the scalp fields, does not generally have an unique solution [24] (i.e, different dipole arrangements could lead to similar fields at the scalp). State of the art methods use a combined spatio-temporal analysis of the fields and a set of physiological constraints to find accurate and robust solutions [83]; those methods are divided in two families: parametric and imaging techniques [83]. Imaging techniques place equivalent dipoles all along the cortical grid with different fixed locations and orientations, and then they adjust each dipole magnitude at each time instant in order to maximise the similarity between the measured fields and the fields that such configuration would elicit in the scalp [83]. Dipole fits derived using imaging reconstruct the observed fields with a greater accuracy but, in exchange, their results can be difficult to interpret [83].

Parametric techniques simultaneously fit the position, orientation and magnitude of a much more reduced number of dipoles [83]. Parametric methods require to specify the number of dipoles, as well as their prior location and orientation, but the results offer a more direct physiological interpretation of the field sources.

In auditory experiments, dipole fitting is generally performed over the averaged elicited fields across 100-200 trials using a parametric approach assuming a single dipole in each hemisphere. In these cases, the solution is usually unique and numerically stable [83]; the dipole model is generally accepted if it can explain more than 90% of the variance of the data. Multi-dipole analyses, placing 2–3 dipoles in each hemisphere, are less common but also frequent in auditory experiments.

Differences between EEG and MEG

Evoked fields at the human scalp are typically of the order of 10^{-12} Tesla; in comparison, the earth magnetic field intensity is of about 10^{-5} T. MEG sensors, called *SQUID*, are based on Josephson junctions that exploit the properties of superconductivity in order to detect subtle changes in the magnetic field at the scalp [24]. SQUID sensors are held in a cryogenic storage dewar filled with liquid helium that keeps the sensors within the superconductivity temperature regime at around 4 K. Moreover, the MEG machinery needs to be electromagnetically isolated in order to avoid interferences from electric devices nearby [24]. These constraints make MEG facilities rare and expensive.

In comparison, EEG is extremely cheap: a reasonably good EEG equipment can be purchased for less than £3000 in the market. In addition, electric fields in the scalp are relatively strong and thus recordings can be taken in a non-shielded room [84].

However, MEG is better suited for auditory experimentation for a number of reasons: 1) MEG preparation times are around 30 minutes, whilst EEG requires manually placing the electrodes on the subject; 2) MEG state of the art machines present up to 306 SQUID sensors, offering a much larger resolution than typical 32/64 channels EEG sets (modern equipments can reach up to 264 channels, but price and preparation times for those sets are prohibitive) [83]; 3) MEG sensors are blind to radial fields [24], which naturally filters out activity originated in adjacent regions of auditory cortex that are not relevant to pitch-related experiments; 4) the scalp is transparent to magnetic fields but not to electric fields, which allows MEG to provide a better signal-to-noise ratio than EEG in response to cortical activity [82]; 5) EEG auditory responses are only prominent along the mid-line electrodes, failing to detect hemispheric asymmetries [83]; 6) EEG electrodes are phase dependent, which affects dipole source modelling decreasing the accuracy the dipole localisation [82].

However, magnetic fields elicited by subcortical activity vanish at the scalp, and only EEG techniques can be used to investigate activity at this level. Moreover, EEG is essential to capture potentially important radial sources from auditory cortex and neighbouring areas [85], although EEG experiments show that such radial sources do not seem to correlate with perceptual features [85]. In any case, EEG and MEG are not mutually exclusive, and some experimental sets combine EEG with MEG to simultaneously record subcortical and cortical activity from different areas in order to investigate functional communications along the auditory pathway [85].

Auditory evoked fields/potentials

The averaged gradient of the magnetic field elicited in human auditory cortex after the presentation of an auditory stimulus, called *auditory evoked field* (AEF), displays a stereotypical trend consisting of set of consecutive onset transients followed by a sustained response that lasts until the stimulus' offset [83] (see Figure 2.8). The electric averaged fields are known as *auditory evoked potentials* (AEP) and show a similar stereotypical trend. In this section, we will discuss mainly MEG results.

The frequency following response

The *frequency following response* (FFR) is an auditory evoked potential elicited in brainstem that, as a result of the phase-locking of the neural activity in the early stages of the auditory nerve, preserves the spectral content of the stimulus' waveform [86, 87]. Since MEG recordings are blind to subcortical sources, the FFR is typically measured using EEG sensors [87].

A cortical equivalent of the FFR has been reported in a recent MEG study using low-frequency speech-related stimuli [88].

Cortical dynamics of the evoked fields

The first transient is elicited around 19 ms after stimulus' onset and peaks at ∼30 ms [89]; it is called *P30*[1]. The P30 is followed by a slightly larger transient known as the *P50*. Depending on the number of averaged trials, these two early components are not always independently resolved.

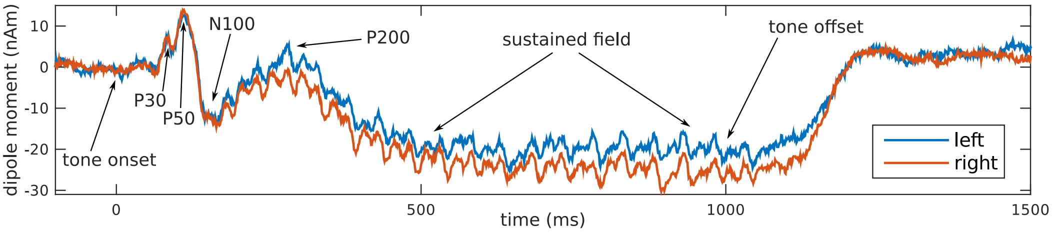


Figure 2.8: Typical notation in auditory evoked fields. This particular field is the dipole moment elicited in alHG by a succession of 20 ramped sinuoids, averaged across 27 subjects and 120 trials [1].

Around 100 ms after tone onset, the fields show a third large deflection called *N100*, followed by a last positive transient, termed *P200* [83]. After the P200, the field asymptotically converges to a negative steady state, known as the *sustained field*, that is held until the offset of the tone, when the polarity of the field rises again up to the original baseline [83] (see Figure 2.8).

The different transients and the sustained field are observed in general locations of Heschl's gyrus and planum temporale, but the exact location of the dipole sources around each of the peaks is slightly different [83]. Moreover, the properties of each transient (i.e, exact latency and depth) are sensitive to different parameters of the stimuli, indicating that each of them reflects a different stage of cortical auditory processing [83].

[1] This is standard electrophysiological notation: *P* stands for positive (negative transients are termed *N*-) and 30 stands for the typical latency. Sometimes, MEG transients have an *m* suffix to differentiate them from their EEG counterparts.

The P50 complex

The P50 complex consists of three transients known as *N19*, *P30* and *P50*. The N19 is the earliest response elicited in auditory cortex after stimulus onset and shows a higher amplitude for high-frequency than for low-frequency sounds [90]. The P30 and the N19 seem to be elicited in the same area of the auditory cortex, but the P50 dipole is located in a separate location near the generators of the N100 and the sustained field [90].

Experiments using up-chirps, frequency modulated stimuli that compensate for the relative delay observed in the basilar membrane when processing low tones with respect to high tones, revealed that the P19-P30 generator might be responsible for integrating inputs incoming from different cochlear channels [89].

The P50 latency directly scales with the period of the stimulus, in accordance with temporal shifts observed as well in lower subcortical structures [91].

The different sources of the N100

The N100 typically peaks at around ∼80–130 ms after stimulus onset [92]. Unlike earlier components of the AEF, the N100 is the result of several aggregated transients sourced at different locations of auditory cortex showing different integration times [93]. Although EEG experiments spotted up to 6 independent sources of the N100 [92], MEG recordings filter our the radial generators and typically reveal only 2–3 tangential sources [22, 94]. Interestingly, the properties of each of the tangential sources are strongly correlated with one of the three perceptual dimensions of single tones: pitch [22], loudness [95], and timbre [94].

Unlike earlier transients, the N100 amplitude shows a certain adaptation to stimuli: repetitions of the same tone elicit lower amplitudes in consecutive repetitions [96, 97]. In melodic contexts, a shift of the N100 generator towards more anterior locations is observed together with the attenuation effect, indicating that the adaptation might be partially caused by modulatory effects from higher cognitive areas [96].

Moreover, the N100 depth seems to be moderately affected by attention processes [98].

The energy onset response. Early studies in EEG already show a strong dependence on the N100 amplitude and latency with stimulus intensity [95] (see Figure 2.9). This correlation seems to be an effect of cortical integration: louder stimuli elicit a stronger response at the hair cells that is reflected in the cortical activity. Moreover, since stronger signals show a larger signal-to-noise ratio, evidence is accumulated faster, explaining the latency dependence. This hypothesis is further supported by the correlation found between the N100 amplitude and the stimulus duration in short tones [93].

More detailed MEG experimental paradigms found that only one of the N100 tangential sources, termed *energy onset response* (EOR), shows sensitivity to loudness [22]. The EOR can be isolated from the other subcomponents of the N100 by using stimuli not eliciting a specific pitch, such as white noises [22]. Experiments with noises located the EOR source in planum temporale, adjacent to Heschl's gyrus [22].

The pitch onset response. Early experiments show that the N100 latency and depth also show a certain sensitivity to the perceived pitch: lower frequencies elicit an N100 with a larger amplitude [91] and later latency than higher pitched tones [22, 23, 91, 99]. Moreover, sounds eliciting a stronger pitch sensation also elicit a deeper N100 [22], in agreement with the correlation between salience and activation in HG observed in fMRI studies [19].

A more detailed study of the N100 behaviour reveals that only one of the subcomponents of the N100, termed *pitch onset response* (POR) in analogy to the energy onset response, shows sensitivity to pitch. The POR source is located in the anterolateral section of HG (alHG) [22, 23], in agreement with fMRI and anatomical results that found pitch-selective activation in the this same area of auditory cortex [19, 50, 54, 65] (see Figure 2.11).

Since the EOR and the POR dipoles are too close to each other to be simultaneously resolved using two different dipoles, the two subcomponents are separately studied using iterated rippled noises preceded by a white noise with the same power spectrum [22]. The noise onset first triggers an EOR, systematically peaking at 100 ms after onset; afterwards, if the transition between the noise and the IRN is smooth, the onset of the IRN elicits an isolated POR [22]. IRN parameters (the number of iterations, which controls the pitch strength; and the delay, that controls the pitch value) can be modified to study the behaviour of the POR (see Figure 2.10). The depth of the POR depends on the strength of the elicited pitch, as shown in Figure 2.10b; the latency of the component depends linearly with the period T of the elicited pitch with factor four: Lat $= 120\,\text{ms} + 4T$ [22], see Figure 2.10c. This dependence is four times greater than the factor of the relationship found in the P30's latency [91].

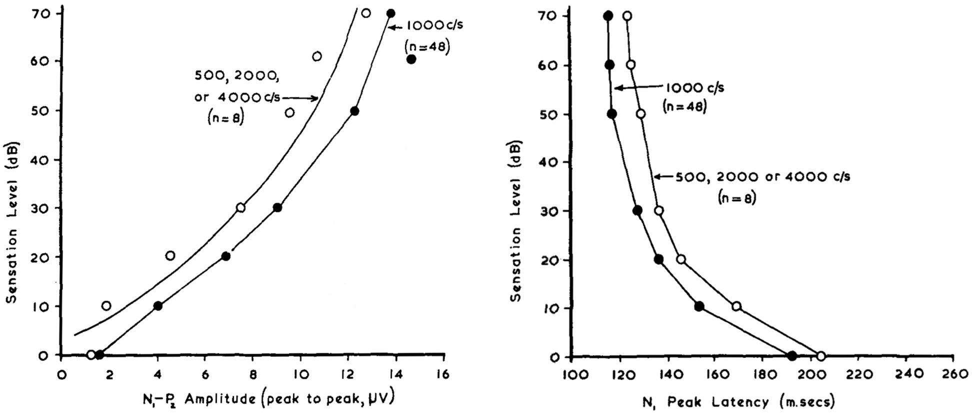


Figure 2.9: Correlation between the N100 depth and latency with the perceived loudness. Depth is portrayed in the left as the N100-P200 peak-to-peak distance; latency is displayed in the right panel. The specific latency and amplitude dependence with frequency described as *cycles per second* (c/s) in the figure has not been experimentally replicated. Figure from [95], Fig. 4.

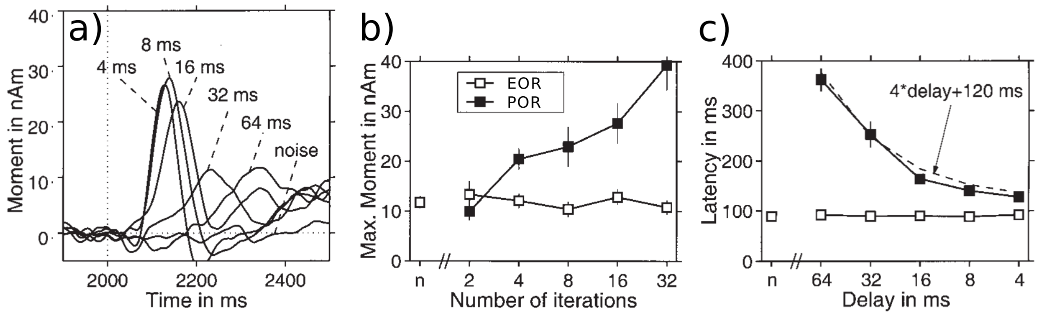


Figure 2.10: N100's latency and depth dependence with pitch. Dipole moments of the POR and EOR elicited using white noises and iterated rippled noises. a) Waveforms around the transition (2 seconds after the onset of the white noise) for IRNs with different delays. b) POR depth dependence with the number of iterations of the IRN. c) POR latency dependence with the delay of the IRN. Figure adapted from [22], Fig. 4.

Timbre. In parallel with the EOR and POR, a third subcomponent of the N100, termed here the *timbre onset response* (TOR), can be studied using an experimental paradigm analogous to the noise-to-IRN setup. An isolated TOR can be triggered by concatenating two tones with identical pitch and loudness but eliciting different timbre sensations [94]. The TOR generator was found to be located in planum temporale [83, 94].

Spectral complexity can also affect the N100 latency: for instance, HCTs elicit later latencies than pure tones [91]. Whether this effect is related with timbre processing or a result of pitch processing mechanisms is, however, still unclear [91].

The P200

The P200, peaking at ∼150–250 ms, is the latest transient observed in the AEF before the sustained field [100]. The P200 overall generator is located in planum temporale [22], although at least two separate sources have been reported in the literature [100].

The P200 latency and amplitude are also sensitive to the stimulus' properties, mimicking the behaviour of the N100: later latencies for low tones, stronger and earlier P200s for louder sounds [100]. This tendency, together with the larger time constant of the transient, seems to indicate that the P200 is elicited by mechanisms at a higher stage of auditory processing and receiving input from the N100 generator [93, 101].

Accordingly, the P200 amplitude is sensitive to musical phrasing context [101] and even to the harmonic context in which the tone is presented [102]. Moreover, the P200's amplitude, but not the N100's [103], is modulated by the musical training and cultural background of the listeners [104].

Sustained field

The sustained field follows the P200 and starts with a negative inflection around 300 ms after tone onset that builds up slowly to a saturation point reached at ∼400 ms after stimulus' onset [83, 105].

Attention, which plays a higher-order cognitive function, has a much larger effect on the depth of the sustained field than it has in the N100 dynamics [98], suggesting that the SF is elicited by processing streams in a higher hierarchical level than the sources of the earlier (P30–50 and N100) transients [91].

Pitch- and Energy- sustained fields. The magnetic SF is generated by two separated sources analogous to the EOR and the POR: a posterior source, located in planum temporale and sensitive to loudness, here called the energy-related sustained field (ESF); and a more anterior source, located in Heschl's gyrus, that arises only when the stimulus elicits a pitch sensation [105], here called the pitch-related sustained field (PSF) (see Figure 2.11).

The reset of the AEFs and the offset delay. The sustained field is interrupted when a drastic change is introduced in the auditory input. If the new stimulus has the same energy but different pitch as the previous one, new POR and P200 transients are elicited and a new pitch-related sustained field arises [107]. If the new stimulus presents different loudness, it also triggers new P30, P50 and EOR transients.

The sustained field is also interrupted when the auditory input stops leading to a period of silence. After the silence onset, the SF holds during a short time called *offset delay* that, in the case of the pitch-related SF, is roughly equal to twice the period of the perceived pitch [105]. If the same stimulus is presented within the offset delay, the early transients are not triggered and the sustained field continues with no sudden interruptions [107]. If the period of silence exceeds the offset delay, the onset transients are triggered once again [107], indicating that the SF is performing a sort of integrative process triggered by the N100-P200

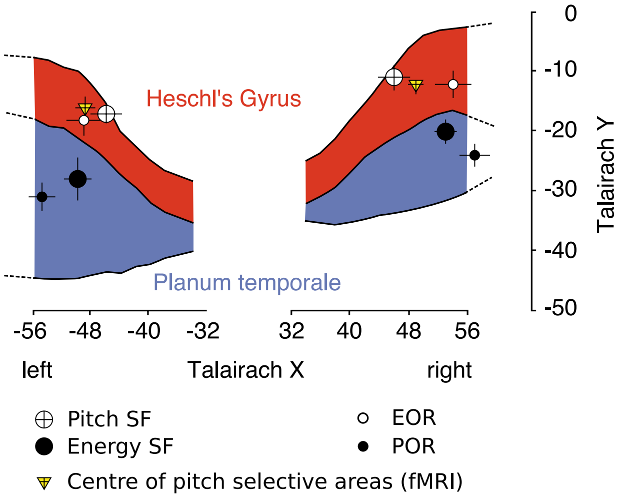


Figure 2.11: Cortical location of the POR, EOR and their associated sustained fields. Figure adapted from [106], Fig. 4.

complex [107]. Accordingly, listeners report an unique stimulus with a silent gap when the silence lasts less than the delay offset, and two separate stimuli separated by a silence when the silence is larger than the delay offset [107].

Global effects

Intensity. Some phenomena affect most of the constituents of the auditory evoked fields in a similar fashion. For instance, an increased sound intensity results in generally stronger fields [95]; this property is often exploited to increase the signal-to-noise ratio in MEG experiments, where stimuli are typically delivered at around 70–80 dB (SPL).

Habituation. Habituation, a progressive decay in the intensity of the responses observed in successive repetitions of the stimuli, is also generally observed in the N100-P200 complex [100]. Short term adaptation is observed in the N100-P200 amplitudes for inter-stimuli intervals (ISI) shorter than 10 s, but rather than being a progressive decay, it seems to affect only the second repetition of the stimulus [97].

Long-term habituation has a slower effect, which reaches its maximum around 30 minutes after the beginning of the experiment [108]. In order to avoid potential biasing effects produced by long-term habituation, different conditions of the stimuli are often uniformly distributed along the experimentation.

Lateralisation. Effects discussed so far in this section are observed in both cortical hemispheres. However, high order auditory functions seem to show a certain hemispheric specialisation: right hemisphere responses are typically stronger during music processing, whilst left hemisphere responses are generally stronger during speech perception [32].

The *asymmetric sampling in time* (AST) theory [33] explains this phenomenon as a hemispheric specialisation in temporal scales: the right hemisphere is suggested to respond preferably to processes requiring longer time scales, whilst the left hemisphere responds preferably to short modulations [33].

Hemispheric asymmetry is also observed in the N100 and the sustained field during the processing of rapidly modulated stimuli [1].

Musical training. FFR elicited by musicians seems to preserve the stimulus spectra with a greater fidelity than the FFR elicited by non-musicians. Subjects with musical experience also present a Heschl's gyrus twice as big as the average listener [109]. This size difference is reflected in a significant increase of the P30's amplitude, but not in later responses such as the N100 [109], although larger P200 responses in musicians have been reported in the literature [110]. Moreover, the N100-to-P200 peak response to a given stimulus increases when the subject is specifically trained to recognise such stimuli [111].

Discussion

The phenomenology of pitch perception

Pitch is elicited in human listeners by periodic stimuli with fundamental oscillatory frequencies between 30 Hz [77] and 4000–5000 Hz [41]. These limits are in line with those of the estimated phase-locked activity in the auditory nerve [40].

A pitch sensation can be triggered by almost any stimuli holding a periodic element, no matter its spectral shape: pure tones, harmonic complexes, iterated rippled noises, click trains; they all elicit a clear pitch sensation [5]. Variations in loudness and timbre have no effects over the perceived pitch value [70], although pitch comparison performances can be affected by strong timbre differences across stimuli [112].

Generally, the pitch sensation elicited by a tone with a fundamental oscillatory frequency f_0 is equivalent to the pitch elicited by a sinusoid with frequency $f = f_0$ [5], but alternated-phase harmonic complex tones with a missing fundamental can elicit a $2\,f_0$ if its harmonics are not independently resolved in the cochlea [60].

Neural representation of pitch along the auditory pathway

The neural representation of pitch varies along the stages of the ascending auditory pathway. At the auditory nerve, pitch-relevant information is coded in the phase-locked activity of the different cochlear channels of the auditory nerve [43]. Two different representations coexist at this stage: a time-code, represented in the temporal structure of the neural activity in the cochlear channels; and a place-code, represented in the overall firing rate of each of the cochlear channels [43].

In the inferior colliculus, we can still observe the tonotopic spectral arrangement [14, 46] and the fine temporal structure of the phase-locked neural activity generated in the cochlea [15]. A third, periodotopic, representation, reflecting a spectral analysis of the phase-locked activity at each channel, arises at some intermediate stage between cochlear nucleus and the inferior colliculus [19, 50] and is widely present in the latter [15, 47].

In primary auditory cortex, phase-locked activity vanishes over 50–200 Hz [16]. Tonotopy [54, 59], and probably periodotopy [46, 57], are observed in Heschl's gyrus and the rostral section of AC.

Moreover, a set of neurons in the low-frequency tonotopic region of HG respond selectively to pitch [19, 50, 54, 65]. This region is often identified as the *pitch centre* of auditory cortex [57]. The internal organisation of the region has not been explored in humans, but mammal studies seem to indicate that groups of more than 10 neurons are collectively selective to different pitch values [68].

Mechanisms underlying the transformations between neural representations

The temporal representation of pitch-relevant information in the early auditory nerve is a consequence of the synchronisation between the basilar membrane and the stimulus waveform, that give rise to the phase-locked neural activity observed along the subcortical auditory pathway [5, 15].

The spectral tonotopic representation found in the subcortical pathway and in primary auditory cortex is a consequence of the stiffness gradient along the basilar membrane that enforces different locations of the membrane to respond selectively to different spectral ranges [5, 43].

The rate-code representation of temporal information found in inferior colliculus (i.e, periodotopy) and putatively in A1, seems to arise between the cochlear nucleus and the IC [19]. Although the specific mechanisms underlying this transformation are still unclear, some theories of how this processing is carried out will be explored in the next chapter.

How the pitch-selective representation observed in the putative pitch centre in HG arises is still unknown; this problem will be addressed later on in this thesis.

Evoked field dynamics during cortical pitch processing

Among the plethora described above, only two components of the auditory evoked fields above seem to reflect pitch processing in cortex: the POR component of the N100 [22,91,99], and the pitch-related sustained field [105–107]. Cortical generators of the POR and the PSF are located in distinct but adjacent places of anterolateral Heschl's gyrus [22, 105], near to the location of the putative pitch centre reported in fMRI studies [19, 50, 54, 65] (see Figure 2.11).

The POR latency scales with four times the period of the stimuli, indicating that cortical generators need to integrate at least four cycles of the stimulus' periodic structure to robustly extract the elicited pitch value [22]. Accordingly, psychophysical experiments using short tones report that robust pitch identification is only possible for durations of over four repetition cycles [43].

Similarly, the delay observed between the stimulus' offset and the pitch-related sustained field offset shows a dependency on twice the period of the stimulus [105], indicating that the PSF cortical generators need to wait for two repetition cycles to confirm that the stimulus is no longer present. Accordingly, psychophysics show that silent gaps under two repetition cycles are identified as a stimulus imperfections, whilst silent gaps over two repetition cycles are identified as a silence period separating two distinct instances of the stimulus [105].

Functional organisation of human auditory cortex

Generators of the EOR and ESF are located in adjacent areas of planum temporale [22,105]; the POR and the PSF are located in more anterior positions in alHG [22, 105]; and the timbre onset response is located in a distinct area of planum temporale. This topological arrangement seems to indicate that there are at least three distinct cortical mechanisms underlying the perception of each of the three auditory dimensions.

A more subtle topological organisation seems to underlie processing within each perceptual dimension, as suggested by the spatial separation between the sustained fields and the onset responses [105]. Accordingly, attention has different effects over the fields evoked in different sections of auditory cortex: it has no effect on the activity at pmHG (equivalent of the core in mammals), a subtle effect on the N100 elicited in alHG (equivalent of the belt in mammals), and a strong effect on the sustained field, elicited near the N100 generator [98]. In addition to the bottom-up/top-down active connections found between the two sections of HG during pitch processing [26], these findings suggests that pitch is extracted by a hierarchy of cortical processing centres [52]. Together with higher level cognitive areas, this hierarchical organisation might play a crucial role in the contextual processing of auditory objects.

Chapter 3

Subcortical processing and abstract cortical models

The different processing stages of the auditory system are complex, highly non-linear systems, whose behaviour escapes analytical formulations. Thus, auditory theories rather rely on models designed to approximately simulate specific aspects of auditory processing. In this chapter, we will review the most relevant theories of pitch processing, from the auditory periphery to central levels, introducing crucial results, which will be used to contextualise our cortical model, and arguing the need for yet another specimen within the heterogeneous zoo of auditory models.

The auditory periphery

Pressure waves arriving in the ear undergo through a series of non-linear transformations along the outer, middle, and inner ear, that are usually modelled separately. A comprehensive model of the periphery typically presents up to five consecutive transformations accounting for: 1) the acoustic effects of the ear canal; 2) the transmission of the tympanic vibration through the middle ear ossicles to the cochlea; 3) the induction of vibrations in the basilar membrane in response to the ossicle movement; 4) the activation of the hair cells in the organ of Corti; and 5) the phase-locked spike trains evoked in the auditory nerve [36] (see Figure 2.1). In this section, we will overview the classical modelling approaches of each one of this stages. Models can be roughly divided in two main families [36]: phenomenological models, that describe the transformations as a set of filtering operations; and detailed physiological models, that model the biophysical mechanics underlying such transformations. Phenomenological models are computationally inexpensive and are often used to reproduce different behaviours of the peripheral system; by contrast, biophysical models can be computationally costly, but they are necessary to study the biological function of the detailed features of the the peripheral anatomy [36].

In this section, we will overview the state of the art in phenomenological approaches; see [36] for a comprehensive review, including biophysical models.

Outer and middle ear

The geometry of the ear canal has a modulatory effect on the incoming waveforms that can be modelled as a cascade of linear filters [36]. Filtering is usually performed in the spectral domain through a transfer function known as the *head-related transfer function*

(HRTF) [36]. Filter parameters are adjusted according to the acoustic transformations observed after placing a microphone inside the ear canal [36]. HRTF parameters vary across subjects, specially for frequencies above 4 kH, and between the right and left ear [36].

The middle ear ossicles transmit the pressure variations elicited in the tympanic membrane all the way to the oval window in the cochlea [5]. This transformation is approximately linear for sound levels under 130 dB (SPL) [36]. Classical phenomenological approaches to describe the effect of the ossicle transmission consist on analogical electric circuits displaying a linear behaviour, whilst modern approaches use a continuous transfer function modelled as a cascade of digital filters [36, 113].

Inner ear and the basilar membrane

The inner ear tranforms the basilar membrane motion into to pressure variations in the oval window [36]. Although cochlear responses are linear when analysed post-mortem, non-linear active effects are found *in vivo* [36].

The frequency-selective behaviour of the different locations of the basilar membrane (see §2.1.1.3) are usually modelled as a bank of overlapping bandpass filters with a monotonically increasing centre frequency [114]. Early versions like the *gammatone* used symmetric and linear bandpass filters; however, the actual behaviour of the cochlear is neither symmetric nor linear: the best frequency of a cochlear channel (i.e, the stimulation frequency at which the channel shows the largest activation) is lower than the average frequency of the response curve, and the bandwidths depend on the intensity level. This complex behaviour was implemented in the *gammachirp* [115], which consisted of an asymmetric filterbank followed by a cascade of level dependent high- and low- pass filters accounting for non-linear effects [36].

Composite models use a symmetric gammatone-like filterbank in parallel with a second, *control path*, that modulates the time constant of the bandpass filters according to the sound's intensity level [36]. The time constant of the filters regulates their gain and bandwidth, thus reproducing active cochlear effects [36].

The dual-resonance non-linear filter (DRNL) uses two parallel asymmetric filterbanks to reproduce the non-linear and asymmetric response of the BM [116]. The first filterbank presents a narrow bandwidth and non-linear responses; the second one presents a broad bandwidth and linear responses and shifted centre frequencies with respect to the first bank [116]. The model's output accounts for active processing by continuously weighting the contributions of the two filterbanks according to the sound's intensity: the first bank is most prominently active under the low-level regime, the second one dominates the dynamics in the high-level regime [116].

A recent model by Zilany and colleagues [39, 113, 117] combines the techniques of composite models and the DRNL in order to reproduce the behaviour of the BM in a more precise way. The model presents a similar structure as that of the DRNL, but the first, wide filterbank, is here modulated by a parallel control path similar to that of the composite models (see Figure 3.1).

Organ of Corti and hair cells

Hair cells in the organ of Corti transduce the mechanical displacement of the BM into electric potentials, that release neurotransmitters ultimately responsible for the auditory nerve activity [113]. Hair cells respond only to one direction of the BM displacement and thus are often modelled as half-wave rectifiers [36].

Besides the frequency-following spike trains, the hair cells present a DC component attributed to the resistor-capacitance properties of the hair cells membrane that deteriorates the phase-locked fidelity of the neural activity [36]. The DC-to-AC ratio increases with the

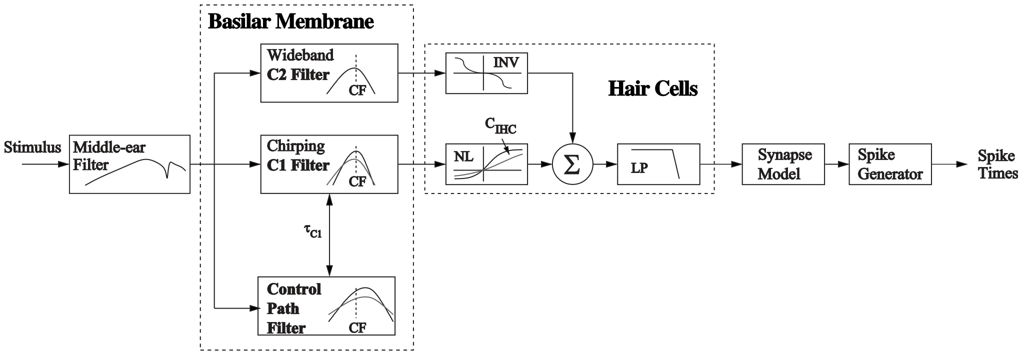


Figure 3.1: Simplified scheme of Zilani's model of the auditory periphery. Figure adapted from [117], Fig. 2.

oscillation frequency, explaining the loss of phase-locking over a certain limit frequency [36] (see Sect. 2.1.1.2). The overall effect is often modelled as a lowpass filter applied after the half-wave rectification [36].

In Zilani's model, different hair-cell transduction processes are applied over the wide- and narrow- filterbank outputs; the DC component is only applied over the high-frequency channels of the wide gammachirp [113].

Auditory nerve activity

Hair cell's electric potentials release glutamate into the synaptic interface between the hair cell body and the auditory nerve dendrites, generating the earliest neural activity signal in the ascending auditory pathway [3]. Glutamate release is modelled as a stochastic process, whose instantaneous probability is related to the hair cell potential by means of a monotonically increasing function known as the *synaptic gain* [117]. Synaptic gain parameters can depend on the centre frequency of each cochlear channel [117].

The release probability is also constrained by the availability of the neurotransmitter in the presynaptic area [36]. After a series of intensive release events, the reservoir of glutamate drops off, causing a decrease in the spike rate at the auditory nerve; this phenomenon is known as *adaptation* [36]. Up to three different adaptation time constants have been observed in the auditory nerve [36]. Adaptation is modelled using a cascade of three reservoirs of neurotransmitters feeding each other in a hierarchical way, and presenting different replenishing time constants [36]. Besides the exponential adaptation resulting from the reservoir dynamics, further short- and long-term adaptation effects, following power-law dynamics, are observed in the auditory nerve [117]. Zilany's model considers power-law adaptation in a phenomenological way, on top of the reservoir effects [39, 117] .

After spiking, neurons show a relaxation time of $\sim 330\,\text{ms}$ during which a second spike cannot be triggered: this is known as the *absolute refractory period* of the cell [13]. More realistic models of the auditory nerve include the effect of refractory periods for greater accuracy [36].

Top-down efferents

Several components of the peripheral system can be actively modulated by top-down efferents originated in the auditory pathway [36]. Although these kind of modulatory effects are poorly understood, a model incorporating top-down modulation of the cochlea showed that the efferent system can regulate the firing rate of selective tonotopic regions of the auditory nerve [118].

Subcortical models of pitch processing

Models of pitch processing have been traditionally divided in two families: spectral models, that use the spectral code of the cochlear channels to predict the perceived pitch value; and temporal models, that use the temporal code of the phase-locked activity in the auditory nerve instead [5, 119]. These, perhaps ad-hoc, simplifications are, however, under close review [120]. Dual theories suggest that both, spectral and temporal codes, are used during pitch decoding [5, 119]. In this section, we will first briefly introduce the spectral models, to then focus on temporal models combining phase-locked activity across different cochlear channels in order to extract a robust subcortical representation of pitch.

Spectral models

Pitch phenomenology is traditionally described on the basis of two different pitch percepts: *spectral* or *periodicity pitch*, evoked by the periodic components of pure tones and harmonic complex tones whose harmonics are distinctly resolved in the cochlea; and *residue pitch*, encompassing the remaining effects, ranging from iterated rippled noises to the pitch evoked by HCTs with unresolved harmonics and missing fundamentals [121] (see §2.2.3.1). Spectral theories aim to explain the mechanisms underlying periodicity pitch, suggesting that a different process is responsible for residue pitch.

Whilst PTs elicit a higher activation in the cochlear channels with centre frequencies near the frequency of the tone, HCTs elicit activation across several cochlear channels (see Figures 2.3A and B). Modern spectral models propose that a set of *harmonic templates* map different activation patterns into a single pitch representation [119]. A biologically plausible model of the early stages of the auditory system suggest that connectivity patterns resulting in such harmonic templates would naturally arise in the plastic human brain after a long exposure to different kinds of sounds [121].

Harmonic templates can successfully predict the perceived periodicity pitch in single tones [119]. Spectral models extend the harmonic templates to composite tones evoking different simultaneous pitch values using *harmonic sieves*, comprised in a phenomenological model proposing that single-pitch templates are matched one after another against the cochlear spectral input in an iterative way by a connected neural network [122]. After finding a template matching part of the spectral input, the model decodes the corresponding pitch value, subtracts the template from the input, and repeats the matching procedure until the input is exhausted [119].

Autocorrelation models

Pitch as an autocorrelation

Temporal models assume that periodicity and residue pitch are elicited by an unique mechanism, proposing a more general definition of pitch: given a single tone's sound waveform x, its pitch is equivalent to the pitch evoked by a sinusoid with period T, where $T > 0$ is the minimum finite repetition time that maximises the autocorrelation function of the waveform; see Equation (3.1).

$$T = \min_{T>0}\left(\arg\max_{T}\left(r(T)\right)\right), \quad r(t) = \int dt\; x(t)\, x(t-T) \tag{3.1}$$

This definition predicts the pitch value evoked by most stimuli (an exception is the family of alternated-phase harmonic complex tones, that we will discuss shortly) and will be useful to link phenomenology to pitch modelling, as shown below.

General formulation of the autocorrelation models

The *autocorrelation models of pitch* are a family of models suggesting that the auditory system exploits the principles of Equation (3.1) in order to extract the pitch value from phase-locked activity in the auditory nerve [17, 21, 123–125]. The first formulation of the autocorrelation models by Licklider [123] introduced a mechanism based on two operations, delay and multiply, subsequently applied over the phase-locked activity of each cochlear channel.

Delay refers to a systematic delay of the spike trains (since spiking activity in the AN is stochastic, models often use the instantaneous probability of spiking $p_k(t)$ rather than the actual spike trains) in each cochlear channel k by a series of delays δt_l, representing candidate pitch values T_l [17, 123, 124]. *Multiply* is a second operation that compares each of the delayed spike trains $p_k(t - \delta t_l)$ with the original activity $p_k(t)$ [17, 123, 124].

Results of the delay-and-multiply operation are further integrated using a time constant $\tau_A \simeq 2.5\,\mathrm{ms}$ [123]. A later reformulation of autocorrelation by Meddis and colleagues [124] suggested that contributions from the analysis across different cochlear channels should be aggregated into a final representation $A_l(t)$, termed the *summary autocorrelation function* (SACF); see Equation (3.2).

$$\tau_A \dot{A}_l(t) = -A_l(t) + \sum_k p_k(t)\, p_k(t - \delta t_l) \tag{3.2}$$

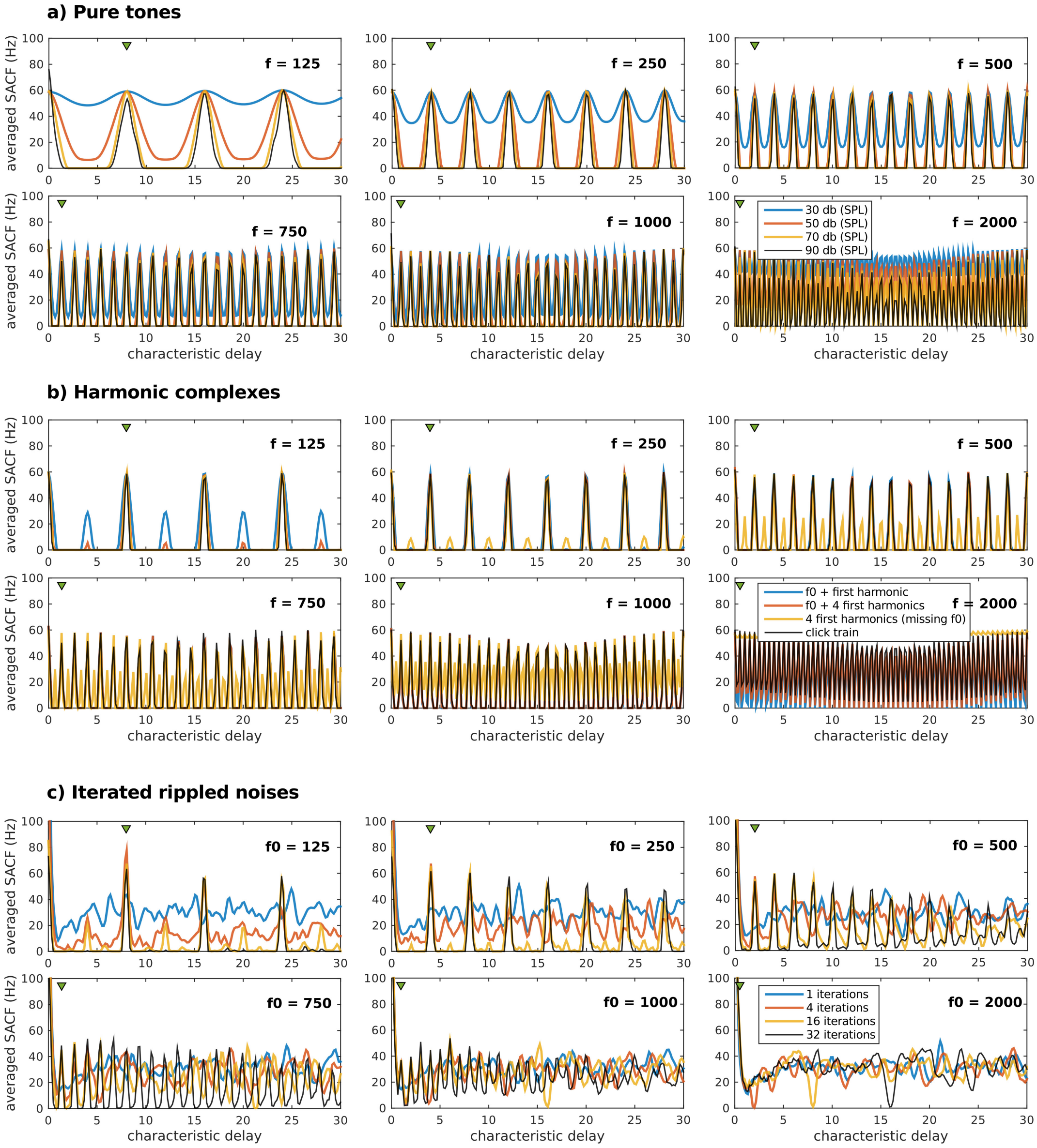
The SACF of the neural activity elicited by a single tone with fundamental periodicity $T = 1/f_0$ presents a series of maxima at $\delta t_l = 0$, and at successive multiples of the tone's period $\delta t_l = T, 2T, 3T, \ldots$; Figure 3.2 depicts the SACF associated to several example tones.

Pitch values and SACF representations

According to Equation (3.1), the evoked pitch corresponds to the smallest non-zero lag δt_l where the SACF presents a maximum; and the strength of the evoked pitch is often related to the activation of the peaks relative to the baseline in the SACF. Selecting this value in the SACF presents, however, several challenges [126]. Placing a heuristic towards low values of the period leads to the peak at $T = 0$, but avoiding the neighbourhood of such a peak would neglect higher pitch values that might be represented in that region [119].

Considering a higher limit of 1600 Hz in the models' frequency range and a bias towards low periods is generally enough to explain the pitch of a wide set of stimuli; however, these heuristics do not work with simultaneous pitch values from composite tones [126]. An alternative strategy is to compute the SACF in non-overlapping groups of cochlear channels, according to their dynamical properties. Using this approach, Balaguer and colleagues [126] accounted for perceptually segregated simultaneous pitches in a phenomenological fashion.

Raw SACF patterns have been successfully used to predict frequency discrimination thresholds [17], indicating that they might be the final representation of pitch in auditory cortex. Under this framework, simultaneous pitch values could be simply represented as overlapping SACF patterns. However, SACF patterns alone cannot be used to explain how listeners judge if a pitch value is higher or lower than another [119].


a) Pure tones

b) Harmonic complexes

c) Iterated rippled noises

Figure 3.2: Averaged SACF associated to different stimuli. The plots picture the averaged SACF (between 100 ms and 200 ms) of the neural activity simulated for different sounds by Zilany's peripheral model [39]. Stimuli are: a) Pure tones with different loudness and frequency. b) Harmonic complexes with different fundamental frequencies and spectral envelopes. c) Iterated rippled noises with different delays $d = 1/f_0$ and number of iterations.

Physiological basis of autocorrelation

The *delay* operation of the first autocorrelation models required a set of delay lines introducing lags up to 33 ms in the auditory nerve activity. These delay lines have not been experimentally observed and lack solid physiological basis [119].

An alternative mechanism, involving chopper neurons in ventral cochlear nucleus and coincidence detectors in inferior colliculus, has been shown to respond in a similar way as the autocorrelation function [17,127]. The alternative model exploits synchronisation properties between the units at CN and IC, that respond selectively to certain periods regulated by the recovery time constant of potassium ions in the chopper neurons [17]. Chopper neurons satisfying the model requirements have been found in CN, making the mechanism much more plausible than its predecessor. Moreover, this idea is coherent with fMRI findings reporting pitch-selective activity arising in CN and IC during pitch processing [19].

The role of cochlear spectral decomposition

Applying the autocorrelation function over each of the cochlear channels is crucial to explain the pitch elicited by the ALT HCTs with a missing fundamental: when the harmonics are individually resolved in different cochlear channels, the autocorrelation function at each channel peaks for lags $\delta t_l = 1/f_0, 2/f_0, 3/f_0, ...$; thus, the lowest period T maximising the autocorrelation function is $T = 1/f_0$, corresponding to the fundamental frequency of the complex. However, if the harmonics are not independently resolved in the cochlea, the entire waveform, that is approximately periodic in $T = 2/f_0$ (see Figures 3.3A and B), is represented in a single channel whose autocorrelation function peaks at $\delta t_l = 2/f_0, 4/f_0, 6/f_0, ...$; thus predicting a pitch of $T = 2/f_0$, according to the psychophysical observations. Averaged SACF for both cases are depicted in Figures 3.3E and F.

Understanding the role of cochlear spectral decomposition leads us to a more accurate definition of pitch: the most repeatedly found period T, as defined in Equation (3.1), across the different spectral bands of the cochlear channels.

Models based on spike coincidences across cochlear channels

Although the autocorrelation models are based on spectral analysis of the phase-locked activity in each separate cochlear channel, the SACF, final output of the model, reports periodicities systematically found across several channels [17]. An alternative neurophysiological implementation of such principles could be based on computing spike coincidences [18] or finding systematic phase-shifts [44] across cochlear channels, and then performing a spectral analysis over the resulting spike trains.

Phase sensitivity in autocorrelation

A major disadvantage of the earlier autocorrelation models of pitch is that they are insensible to the relative phase of the sound's components that are resolved in different cochlear channels [119]. This insensibility is crucial to explain why ALT HCTs with resolved harmonics elicit the same pitch than HCTs with other phase arrangements, but phase interactions are necessary to explain subtle shifts in the perceived pitch values in experiments with shifted HCTs [21,128]. Although extensions of the autocorrelation models can explain these shifts in a phenomenological way [21], they lack biophysical realism. Approaches considering coincidences across cochlear channels allow for across-channel interactions, which allows them to account for such derived phase effects with a greater biophysical realism [18,129].

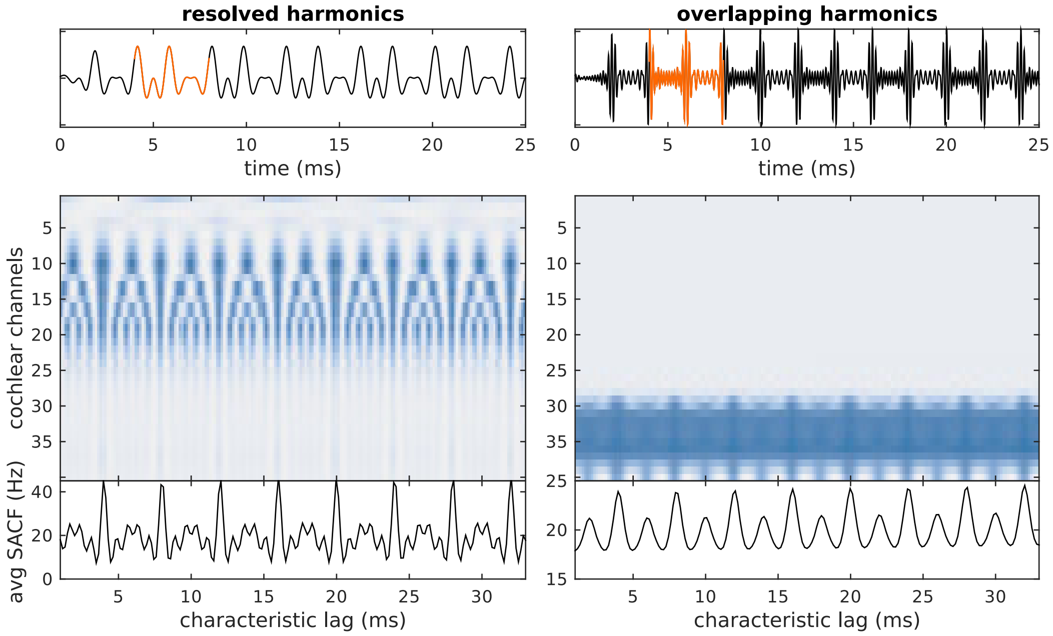


Figure 3.3: Averaged SACF derived for two ALT HCTs with missing fundamentals and same $f_0 = 250\,\text{Hz}$ but eliciting different pitch vales. Top) 50 ms segment of the stimulus waveforms, orange segments comprise a single repetition cycle; a) an ALT HCT with the first six harmonics, b) an ALT HCT with harmonics 15th to 25th. Middle) ACF of each of the 40 cochlear channels considered in the peripheral system [39]: c) harmonics are independently solved across different channels, d) harmonics collude in the high-frequency channels. Bottom) SACF for both stimuli, derived by adding up the individual cochlear contributions, and averaged between 100 ms and 200 ms after stimulus' onset: e) SACF pattern corresponds to a perceived pitch of $T = 4\,\text{ms} = 1/f_0$; f) SACF pattern corresponds to a perceived pitch of $T = 2\,\text{ms} = 1/2\,f_0$.

STDP-based coincidence detectors

One of the models, developed by Erfanian and colleagues [129], suggests that auditory nerve activity from different cochlear channels is analysed by a network of stereotypical neurons whose connections are dynamically adjusted following spike-time-dependent plasticity (STDP)[1]. The neural network receives inputs from a phenomenological model of the auditory periphery, and adjust the neural connections according to STDP rules. After ~ 5000 seconds of learning, the network is tuned to find synchronous activity across cochlear channels, whose inter-spike interval is predictive of the elicited pitch in pure tones and harmonic complex tones with and without missing fundamentals [129].

Slope coincidence detectors

A very recent biophysical model by Huang and Rinzel [18] introduced a more sophisticated neuron receiving phase-locked activity from different cochlear channels [18]. Huang's neuron acts as a slope detector, that activates.14159265359Qq in phase with coincident spikes received across several cochlear channels [18]. As in the previous model, the inter-spike interval of the output reflects the elicited pitch in pure tones and HCTs, but also in click trains and iterated rippled noises [18]. More importantly, the output activity is loudness- and timbre- independent [18], and the neuronal model shows a great biophysical detail.

Neural representations

Huang suggests that the temporal structure of the output of these kind of models might conform the final representation of pitch [18]. Phase-locked spike trains can be used to perform comparisons across tones, and test which of two tones presents a higher pitch value [18].

However, the lack of phase-locked activity in cortex over 200 Hz shows that pitch cannot be represented as a temporal code in auditory cortex [16]. Thus, these models need a later step, also implemented subcortically, transforming the inter-spike-intervals from the spike trains into a rate-place representation.

This process can be carried out by finely tuned oscillators or chopper neurons in cochlear nucleus as those of the autocorrelation models [18]. In any of the two cases, the output of the transformation would yield a similar trend as the SACF patterns: a first peak of activation corresponding to the period of the elicited pitch $\delta t_l = T$, and subsequent peaks on periods characterising the lower harmonics $\delta t_l = 2\,T, 3\,T, 4\,T, \dots$.

Cortical models and adaptive strategies

In this section, we will review three families of models that will help us to contextualise our contribution. Two of those families are based on adaptive pitch processing and present subcortical and cortical stages; the third family describes the mesoscopic dynamics of cortical processing in a general way, not specific to pitch processing.

Adaptive models based on the SACF

Although the SACF itself displays low sensitivity to across-channels phase interactions, stimulus' features affecting pitch strength are also present in fast variations of the SACF that

[1] STPD is a learning framework considering phenomenological rules, first introduced in their modern form to explain sound localisation, that have been widely used since to explain plastic processes in a general way [130]

can be measured using short integration windows [1, 21]. However, longer integration windows are necessary to extract a stable representation of the autocorrelation output elicited by low-frequency tones.

A partial solution is to consider that the integration τ_A in Equation (3.2) is lag-dependent; i.e, that each delay line δt_l is integrated with a different time constant $\tau_A \to \tau_l = 1.25\,\delta t_l$, with a minimum value of $\tau_l \leq 2.5\,\mathrm{ms}$ [125, 131]. In addition, this lag-dependent integration introduces a natural bias towards smaller periods δt_l that accentuates the first peak in the autocorrelation pattern, allowing us to identify the perceived pitch with the largest peak of activation [21, 125]. Some of these models further present a cascade of processes using different temporal integration mechanisms [21, 125, 132].

The generative pitch model and pitch change

The generative pitch model (GPM) addresses the problem of balance between integration and resolution in pitch processing. GPM presents a cascade of integrators that communicate with each other in a bottom-up and top-down manner [21], in agreement with LFP observations about the hierarchical organisation of Heschl's gyrus [26]. GPM is focus on change detection in pitch sequences and pitch gaps, but it is also able to extract, in a phenomenological way, the pitch value of a wide range of challenging stimuli [21].

Top-down efferents allow GPM to dynamically tune the integration constants along the cortical cascade, in order to selectively capture short- or long- term temporal structures in the stimulus' input [21]. The integrators described in this model follow formal neural ensemble dynamics and are plausible in terms of modelling principles [21]. The adaptation of the integration windows in GPM can also be explained in terms of interacting neural ensemble models, but the criteria underlying the adaptive integration was chosen ad-hoc and lacks biological realism [21].

Nevertheless, GPM's adaptive top-down mechanism has been connected with physiological elements of the N100, relating the mean firing rate of the populations encoding the pitch value with the amplitude of the neuromagnetic transient elicited by a wide range of stimuli [1, 21]. The smoothed derivative of the activity at this population has also been found predictive for the POR latency in iterated rippled noises [21].

However, the connection between the GPM activity and the derived fields also lack of biological realism: we would expect the whole mass of neural ensembles to be responsible for the evoked fields [133, 134], rather than a single population whose selection depends on the elicited pitch of the stimulus [21]. Thus, these two correlations seem to reflect phenomenological aspects rather than the actual neural mechanisms underlying pitch processing.

For instance, the peak of the response of the model at the population encoding the stimulus pitch value shows a dependence with pitch strength sourced in the shape of the SACF input [21]. Since the POR depth depends linearly on pitch strength [1, 22], it seems likely that the correlation between the GPM activity and the MEG signal is sourced in this common correlation [1].

The auditory image model

The Auditory Image Model (AIM) introduced by Patterson and colleagues [128,132,135] uses a phase-sensitive variation of the autocorrelation function during the subcortical processing known as *strobed temporal integration* (STI). In STI, the autocorrelation processes in each cochlear channel is substituted by a crosscorrelation between the auditory nerve activity and a train of *strobe pulses*, consisting of a leaky aggregation of the phase-locked spike trains between peaks of activation [135]. Sensitivity to several times scales is achieved by dynamic tuning of the pulse detector threshold [135].

The *stabilised auditory image* (SAI) is the cortical representation of the STI, designed to simulate a highly idealised neural representation of the auditory stimuli, and assumed to underlie the first conscious awareness of a sound [132]. SAI's pitch representation resembles that of the autocorrelation patterns: the characteristic period of the first peak after the zero pole corresponds to the predicted pitch value, whilst the ridge height of the peak is predictive of the perceived pitch strength [136].

Although pulse trains showing similar properties as the strobe pulses of AIM have been observed in alive octopuses [119], this model is often regarded as highly idealised. For instance, there is no physiological candidate mechanism able to implement the adaptive tune of the strobing threshold [135].

MEG studies have qualitatively connected different components of the AEFs with AIM-related processes [89, 108, 137]: the sustained field has been connected with activity in the SAI and the N100 with its derivative [108]. However, these parallelisms reflect an abstract correspondence rather than a functional one: AIM does not attempt to explain how these processes arise in biophysical terms and it fails to perform quantitative predictions of their properties [1].

Dynamic causal modelling

Generative models and DCM

Dynamic causal modelling (DCM) [138–141] is a formal modelling framework designed to analyse mesoscopic (e.g. fMRI, local field potentials, or E/MEG) cortical recordings. DCM assumes that the observed data can be explained by means of mesoscopic interactions between a finite number of cortical regions, receiving stereotypical idealised thalamic inputs [141]. Neural parameters characterising the dynamics of and interaction between cortical regions are tuned using Bayesian optimisation [141].

DCM is often regarded as a generative modelling framework [138]. For instance, In the E/MEG domain, DCM tests candidate configurations by deriving the electromagnetic fields that the candidate dynamics would have elicited [133]. Bayesian optimisation is used to adjust the parameters of the cortical network in order to maximise the fit between the observed and generated fields [133]. Prior positions and orientations of the cortical generators are often calculated via parametric dipole fitting analysis. More sophisticated DCM formulations consider the whole cortical tissue as a continuous neural field [142].

Cortical columns in DCM

DCM cortical networks consist of a number of interconnected cortical columns. Columns are modelled as blocks with three neural ensembles representing: 1) pyramidal excitatory neurons (PE), 2) spiny stellate neurons (SS), and 3) inhibitory interneurons (II) [141]. Each ensemble reflects properties of neurons located in different cortical layers.

Thalamic input is fed directly to the SS ensemble, which further communicates with the II and PE populations [141]. E/MEG fields are derived from the activity of the pyramidal excitatory neurons [133]; the activity of the II and PE ensembles are considered as hidden factors in the model.

DCM cortical columns are organised in larger hierarchical networks. Connections between columns are considered *bottom-up* if they connect a presynaptic PE ensemble with a postsynaptic SS populations and *top-down* if they target PE or II ensembles [143]. The specific number of columns considered in a DCM is a prior parameter of the model.

Within a column, each neural population is modelled according to neural ensemble theory. Population dynamics are based on results from statistical mechanics, deriving the evolution of statistical descriptors of the neural population (e.g. the spiking probability distribution) in aggregations of neurons. Dynamics describing the PE, SS and II ensemble evolutions are often structurally similar, but display different parameters [141].

The Laplace assumption

Parameter fitting in generative models is computationally expensive, since network dynamics have to be simulated in each iteration of the optimisation process. DCMs alleviate the computational cost by adopting the *Laplace assumption*, based on the hypothesis that the spiking probability distribution of the neurons within an ensemble follows a Gaussian distribution [144]. The Laplace assumption simplifies several dynamic properties of the ensembles, making them analytically tractable and notably reducing the computational cost of the simulations [144, 145].

Similarly, synaptic inputs are modelled as either Gaussian processes or as constant values, which further simplifies the network dynamics [145].

Limitations making DCM unfit for our investigation

Although DCM is a powerful tool to establish the hierarchical role of different cortical regions during neural processing, its limitations make it unfitted for our purposes.

A first limitation is grounded in the spatial resolution of E/MEG: spatial scales are too large to represent pitch-selective regions responding to different pitch values in the human audible range. In a study using intracranial LFP [26], which presents a higher signal-to-noise ratio than E/MEG and thus a finer spatial resolution [24], a DCM analysis divided Heschl's gyrus into three regions.

In contrast, a mechanistic explanation of cortical pitch processing would require us to distinctly model the behaviour of pitch-selective populations within alHG, which requires a resolution two orders of magnitude higher [21]. MEG recordings during pitch processing could be modelled considering a single observable, generated using the aggregated activation in alHG, and regarding the dynamics of the individual pitch-selective ensembles as hidden factors. However, DCM considers one observable for each separate cortical column [141].

Moreover, the dependence of the features of the evoked fields with the stimulus' pitch cannot be modelled using DCM, due to the simplicity of the thalamic input considered in this framework. Complex input structures, that would increase the computational cost of the Bayesian optimisation up to intractable levels, are necessary to model the dependency of the observed fields with the stimulus' features.

A case study on cortical processing of pitch strength

In this Section, we will introduce a case study on pitch strength prediction combining modelling and MEG recordings using abstract pitch perception models. These results were extended and published recently by the authors of this thesis [1].

Introduction

Auditory stimuli that display different attack (onset) and decay (offset) times, are said to be *temporally asymmetric* [135]. Ramped and damped stimuli, consisting of a sinusoid multiplied either by a periodically rising (*ramped*) or decaying (*damped*) exponential function [114, 135] (see Fig 3.4), enable us to study temporal asymmetry in a systematic fashion.

In this case study, we used a family of 10 different stimuli, consisting of concatenations of 20 repetitions of either a ramped or a damped sinusoid, modulated with five rise/damp exponentials with different half life times: 0.5 ms, 1 ms, 4 ms, 16 ms, and 32 ms [135]. The

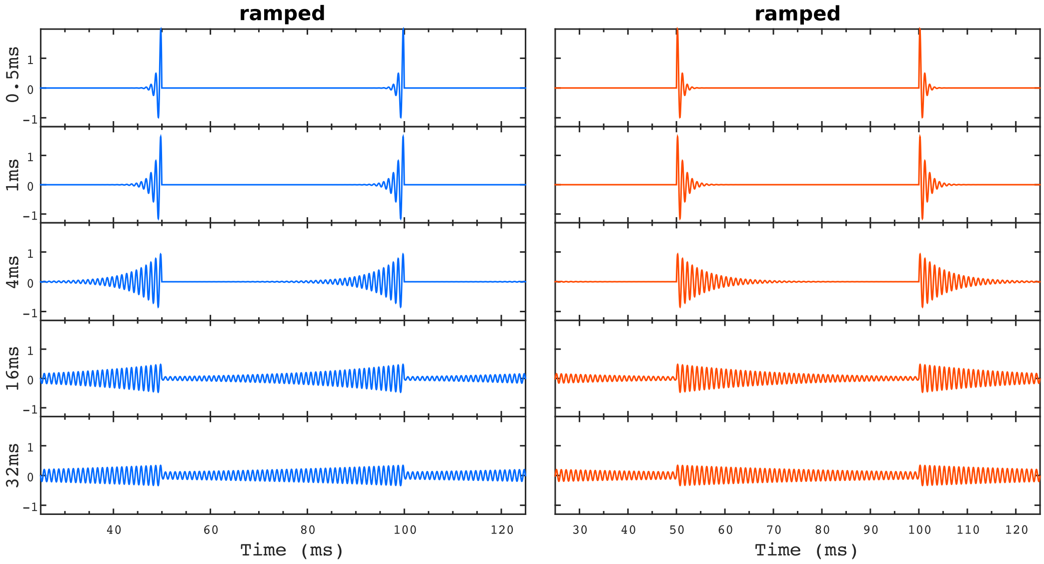


Figure 3.4: Waveforms of ramped and damped sinusoids. Ramped (left) and damped (right) sinusoidal waves with half-life times ($T_{1/2}$) of 0.5, 1, 4, 16, and 32 ms used in the experiment. Note the two periodicities present in the stimuli corresponding to the carrier (1000 Hz) and the repetition period (20 Hz) of the ramped/damped modulation. Figure taken from [1], Fig 1.

concatenations always elicit the pitch of the carrier, set to 1000 Hz in this experiment [135]; however, whilst ramped sounds are perceived as continuous tones, damped sinusoids are perceived as a drumming sound with a lower pitch strength [135].

Ramped and damped sinusoids present identical long-term Fourier spectra; hence, autocorrelation models cannot fully explain such perceptual differences [135]. Here, we will use two models incorporating stimulus-dependent adaptive processing of the auditory nerve activity, the generative pitch model (GPM; see §3.3.2) and the auditory image model (AIM; see §3.3.3, to investigate the perceptual differences between these two families of stimuli [1].

Moreover, we will show that the N100 morphology of the auditory evoked field elicited in anterolateral HG reflects processing of temporal asymmetry in auditory cortex, and that the GPM dynamics are able to account for such dependence [1].

Experimental procedures and modelling approach

Psychoacoustic measurements

We performed psychoacoustic measurements of the pitch strength elicited by each of the 10 different stimuli described above. Modulated sinusoids were presented in a single block of consisting of all possible combinations of pairs of non-identical stimuli (a total of 45 pairs, 90 trials) [1]. In each trial, listeners had to indicate in a two-alternative task without feedback which sound of the pair was perceived as more tonal. After a training session, the block was presented just once. A scale for the relative pitch salience was derived from the results of the paired comparison experiment, using the Bradley-Terry-Luce (BTL) method [146].

Neuromagnetic data

Auditory evoked fields were measured by our experimental colleagues in Heidelberg University. Fields were averaged over an epoch from -500 ms to 1400 ms. A two-dipole model was fitted based on the pooled 16 ms and 32 ms ramped and damped conditions, assuming that the N100 response evoked by all stimuli had the same generators in auditory cortex. Dipole sources were localised in lateral Heschl's gyrus [1].

GPM-derived evoked fields

The first derivative of the GPM output has been shown to be correlated with available neuroimaging data associated to the perception of Iterated Ripple Noises [21]. In the present study, we used a similar approach to compare the dynamics of these predictive units with the morphology of the N100 response evoked by the each of the ramped and damped stimuli.

For each of the 10 stimuli, we matched the response of the model's top layer at the pitch value prediction to the amplitude of the evoked response within a time window of 50 ms surrounding the N100 peak [1]. To fit the peak, we proposed a linear relationship between the amplitude of the model and the amplitude of the MEG signal (see e.g. [142]).

The linear fit was cross-validated across subjects as follows: first, we performed an individual linear fitting for each of the $N = 27$ subjects in the experimentation [1]. Then, parameters of the linear fits were fixed and tested using the evoked fields of the remaining $N - 1$ subjects, yielding to a total of $N(N-1) = 702$ cross-validation folds per stimuli [1].

Results

Figure 3.5 summarises the main results of the case-study: Figure 3.5A shows perceived pitch strength; Figure 3.5B shows the pitch strength predicted by AIM as the mean ridge height of the SAI at the value of the perceived pitch; Figure 3.5C shows the N100 amplitudes; and Figure 3.5D shows the N100 amplitudes predicted by GPM as the maximum of the soft-derivative in a 50 ms window surrounding the N100 peak latency [1].

Pitch strength. Pitch strength increased with modulation half life time values $T_{1/2}$ for both, ramped and damped sounds; moreover, the pitch of the ramped tones was generally judged as more salient than the pitch of their damped counterparts. This difference reached significance for the critical value $T_{1/2} = 4$ ms ($p = 0.0077, n = 13$) and for $T_{1/2} = 1$ ms ($p < 0.001$, $n = 13$) [1].

N100 amplitude. N100 peak amplitude increased with the $T_{1/2}$ of the stimuli for all conditions and was significant for the transition from $T_{1/2} = 1$ ms to higher half-life values (ramped: $p = 0.0003, n = 837$; damped: $p = 0.0039, n = 837$) and the transition from $T_{1/2} =$ 4 ms to higher half life times in the damped case ($p = 0.0146, n = 837$) [1]. Consistently with perceptual results, ramped tones evoked larger N100 than damped ones, with a maximal difference at the critical value of $T_{1/2} = 4$ ms ($p = 0.0008, n = 837$) [1].

We found a high correlation between the magnitudes of N100 and the relative perceived carrier salience for ramped ($R = -0.9597, p = 0.0097$) and damped ($R = -0.9867, p = 0.0018$) stimuli [1].

AIM predictions. Although the trends in Figure 3.5B seem to diverge from the perceptual results in Figure 3.5A, there is a high correlation between the AIM predictions and the measured perceptual trends (ramped: $R = 0.978, p < 0.05$; damped $R = 0.978, p < 0.05$) [1]. The divergence between the absolute values of the trends for ramped and damped sounds can be attributed to non-linearities in the transference function mapping the height of the

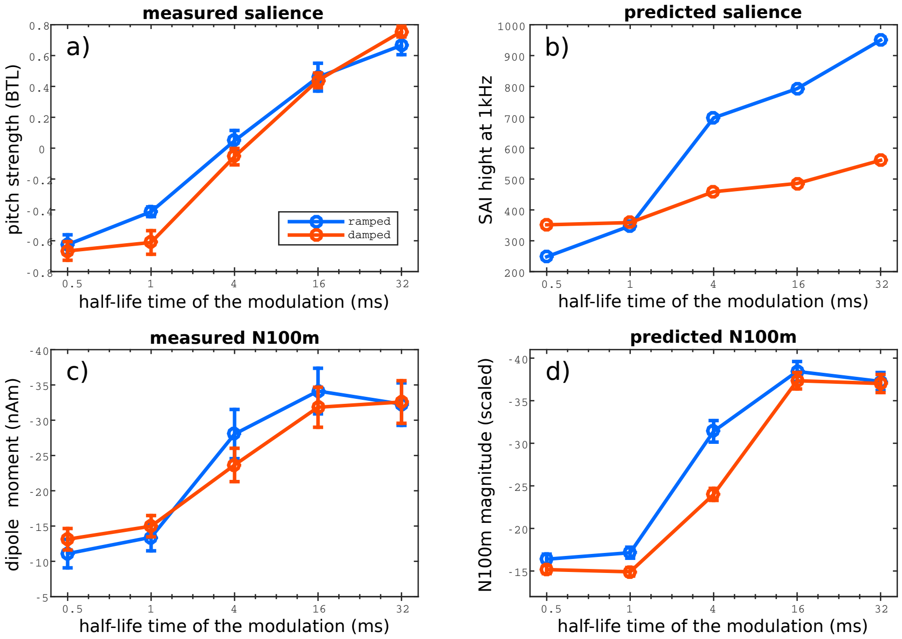


Figure 3.5: Comparison of the perceived salience, N100 magnitude, and the prediction of the two models of pitch. a) Perceived salience estimated by the BTL method and averaged across subjects ($N = 13$). b) SAI mean ridge height at the frequency of the carrier (1 kHz). Ridge height was used to predict the perceived salience of the stimuli [132]. c) Magnitude of the N100 component averaged across subjects. d) Top-down modulated model's predictions for the amplitude of the N100m peak, computed as a linear transform of the derivative of the activation of the top layer population evaluated at the frequency corresponding to the stronger model response. Significant correlations were found between perceived saliency 3.5a) and N100m magnitude (3.5c); between the perceptual observations 3.5a AIM responses (3.5b) and between the N100m magnitude 3.5c) and GPM predictions (3.5d). Error bars represent SME. Figure adapted from [1], Fig 4.

SAI ridge to the perceived pitch strength: although this function has been described as monotonically increasing [136], the exact shape has not been studied in the literature.

Nevertheless, the observed correlation suggest that the strobed integration process effectively extracts the temporal asymmetries responsible for the differences in sensation from the auditory nerve activity [1].

GPM predictions. Differences between model simulations for the N100 amplitude of ramped and damped stimuli were highly significant for a $T_{1/2} = 4\,\text{ms}$ stimulus ($p < 0.0001, n = 702$), consistent with the psychoacoustic and neuromagnetic results [1]. Moreover, modelling predictions show a strong linear correlation with the experimental N100 magnitude for both, ramped ($R = 0.9972, p = 0.0002$) and damped ($R = 0.9899, p = 0.012$) stimuli [1].

Conclusions

Results of this case study, succinctly summarised here, confirm that the alHG sources of the N100 are related to pitch decoding, as frequently reported in the literature (see §2.3.2.4). More importantly, we found that rapid stimulus-adaptive processing is a key element to understand the perception of asymmetric sounds and the observed differences in the N100 morphology [1].

Spectral analysis on the basilar membrane and the neural transduction process enhance temporal asymmetry to a certain extent [147]; however, this enhancement is not sufficient to explain perceptual effects [147]. Classical autocorrelation models [17, 124, 125], although successful in pitch extraction, are unable to extract a faithful representation of temporal asymmetry in the auditory nerve [1]. In contrast, the two idealised adaptive models considered here, successfully amplified this temporal asymmetry and predicted the perceived differences between ramped and damped stimuli.

Furthermore, GPM accurately predicted the magnitude of the evoked N100, suggesting that temporal asymmetry encoding might be mediated by a hierarchical process with top-down driven stimulus-specific integration windows [1].

In summary, our results provide for evidence of the N100 magnitude indicating the presence of a neurophysiological mechanism encoding pitch strength in auditory temporal asymmetry, and suggests that pitch salience asymmetry can only be explained by means of adaptive windows of temporal integration [1].

Although AIM and GPM allow us to understand key aspects of temporal pitch processing, we need to consider a greater level of biological detail in order to unravel the neurophysiological underpinnings of the pitch-related auditory evoked fields.

Discussion

Subcortical mechanisms of spectral processing

Temporal models present a larger prediction power than the spectral counterparts. Although spectral models equipped with harmonic templates and sieves can account for the perception of periodicity pitch, temporal models like the SACF and AIM do not distinguish between periodicity and residue pitch, explaining both phenomena as two expression of a general processing mechanism [119]. However, it should be noted that temporal models do not make use of temporal information alone; on the contrary, the spectral decomposition performed by the cochlea and the propagation of tonotopy along the subcortical pathway plays a crucial role on the processing of complex stimuli (see §3.2.2.5).

Within the temporal models, both autocorrelation and AIM are good candidate mechanisms transforming the phase-locked temporal code into a place representation, akin to the periodotopic arrangement found in inferior colliculus (see §2.4.2). However, the threshold adaptation mechanism of the strobed integration in AIM lacks biological realism [135], whilst autocorrelation is blind to fast changes in the stimulus' waveform and across-channel phase interactions that can affect the perceived pitch strength and provoke subtle pitch value shifts [21]. Moreover, both the SACF and AIM outputs vary with loudness [17, 135].

Loudness dependence can be fixed normalising the SACF representation [148]. Sensitivity to fast changes in the SACF can be addressed in later cortical steps [21], but cortical models describing how this sensitivity is achieved do not attempt to provide a biophysically realistic description of the decoding process.

Huang's slope coincidence detectors provide a mechanistic model to correct for loudness and timbre dependence during the transformation, and further increase the sensitivity to phase interactions [18]. However, this model does not explain how the temporal code is transformed into a periodotopic representation as observed in inferior colliculus or expected in cortex. An autocorrelation-like transformation of the temporal code generated by Huang's neurons would yield a reliable periodotopic code invariant to loudness and timbre; however, a biophysically plausible mechanism in charge of such transformation has yet not been devised.

Neural representation of pitch in the subcortical pathway

Although the underlying mechanisms vary, temporal models described above coincide in the shape of the final representation of pitch derived form the spectral analysis of the phase-locked activity across cochlear channels. The final representation follows a *place rate code*, in the sense that pitch is conveyed by the firing rate (rather than the inter-spike intervals) of different fibres in the auditory nerve.

In the aforementioned models, subcortical afferents are characterised by a characteristic period δt_l, and show a maximum activation when the auditory nerve phase-locked activity presents a periodicity $T = \delta t_l$. A given pitch value f_0 is represented by a harmonic pattern showing peaks of activation at all multiples of its fundamental periodicity, $\delta t_l = n/f_0, \quad n = 0, 1, 2, 3, \dots$; see Figure 3.2. Accordingly, an EEG study exploiting habituation effects found that the neural representations of the pitch sensations evoked by harmonically related IRNs are more similar than the neural representations of non-harmonically related tones [149].

Although harmonic patterns of activation in inferior colliculus's periodotopic axis have not been robustly reported in the literature, harmonic co-activation in frequency-tuned neurons has been observed in a large number of mammal intracranial [62] and human fMRI [59] recordings in cortex.

Pitch models and the auditory evoked fields

POR amplitude The correlation between the POR depth and the perceived pitch strength resembles the correlation between pitch strength and the relative activation of the SACF and SAI peaks with respect to the baseline activity [1, 21, 136]. Thus, the enlargement of the POR amplitude with tonal salience can be explained as a consequence of the increase of the signal-to-noise-ratio in the SACF and the SAI with pitch strength.

POR latency Temporal dynamics of subcortical systems do not explain the dependence of the POR with four times the stimulus's period (see §2.3.2.4). Integration constants in autocorrelation models suggest that enough information to extract a pitch value from inter-spike intervals is gathered after ∼1.25 cycles of the periodic stimulus's. Thus, current subcortical theories of pitch processing are unable to explain the observed dependence.

Cortical models essentially integrate the subcortical input, and they do not introduce additional pitch-dependent delays that could account for the extra ∼2.75 cycles. Moreover, GPM and AIM perform dynamic tunings of the integration time constants, which depend on timbre and loudness variations [21, 135]; thus, they cannot explain why the POR dynamics are exclusively modulated by pitch properties.

Sustained field Although it has been suggested that the sustained field might reflect integrative processes modelled by AIM [108], the SF is elicited around 200–300 ms after tone's onset, whilst AIM starts integrating only a few milliseconds after tone's onset. Cortical models based on autocorrelation are equally unable to explain the late onset of the SF.

Potential mechanisms of cortical pitch processing

Despite the success of temporal models of pitch in explaining psychophysical phenomena, the models reviewed in this chapter fail to explain the specific neural and synaptic mechanisms underlying cortical pitch processing, and how they elicit the associated cortical responses in Heschl's gyrus.

Candidate mechanisms of cortical pitch processing could be based on the transformations mapping the subcortical representation of pitch, here assumed to consist of harmonic patterns of activation as described above, into a receptive-field-like representation where a single neural ensembles activates in response to each of the pitch values represented in cortex.

This transformation echoes the earlier harmonic templates of the spectral models, where cochlear patterns of activation are mapped to single pitch values. However, template matching operating on the subcortical representations considered above presents two advantages: first, unlike in the cochlear template matching, representations evoking a given pitch value present similar shapes [17]; second, the frequency resolution of the SACF for complex stimuli is much narrower than that of the cochlea [119].

The idea of template matching in cortex is also be compatible with the dependence of the POR's latency, and the necessary tone's duration for robust pitch labelling, with the pitch value: integration across several period cycles is necessary to evoke a sufficient number of peaks in the subcortical representation as to make the pattern unequivocally recognisable.

This idea will be explored in the next chapter, where we introduce a novel theory of cortical pitch processing accounting for the onset and sustained dynamics observed in MEG recordings.

Chapter 4

Neural dynamics of cortical pitch processing

Introduction

Biologically realistic theories of pitch often focus on the peripheral processing of input sounds, whilst the specific role that auditory cortex plays in the processing pathway remains a challenge. Models addressing cortical processing (e.g. [21, 132]) often lack sufficient mechanicistic detail. However, a physiological understanding of cortical processing might be crucial to explain the origin of the cortical representation of pitch and the dynamics of the elicited evoked fields.

In this chapter, we will introduce a novel theory of cortical pitch processing describing the mechanisms mapping subcortical harmonic patterns of activation (see §3.5.2) into a stable cortical pitch representation (see §2.4.2). Our candidate mechanism describes the rise of the POR and the pitch-related sustained field, quantitatively predicts the POR's morphology and latency dependence on the tone's elicited pitch, and explains the late onset and offset dynamics of the pitch-related sustained field (see §2.4.4); for the first time to our knowledge.

Our cortical theory is embedded in a comprehensive model comprising peripheral, subcortical and cortical processing.

The model

Overview

The model consists of several processing stages representing different hierarchical levels along the auditory pathway. First, a subcortical array of idealised periodicity detector units, based on the principles of autocorrelation (e.g., [17]), processes the spike trains generated by a state of the art peripheral model [39]. The output of the periodicity detectors is then normalised and used as the cortical model's input as explained below.

The cortical model presents two processing layers, termed here *the decoder* and *the sustainer*. These two networks are putatively located in adjacent locations of antero-lateral Heschl's Gyrus, each one consisting of a population of balanced E/I ensembles (see Figure 4.1). The decoder effectively extracts the pitch value from the subcortical representation, whilst the sustainer integrates the decoder's representation and modulates its functioning through top-down cortico-cortical efferents. This cortical arrangement is reminiscent of the network dynamics observed in other cognitive systems, where perceptual decisions propagate toward higher cortical levels that further modulate the lower-level dynamics [150]).

Patterns of activation generated at the subcortical level present harmonic shapes that peak at frequency values encoding the pitch of the stimuli and all their lower harmonics (see bottom plots in Figure 4.1; more examples are plotted in Figure 3.2; a detailed discussion on the origin of these patterns is provided in §3.5.2).

Each of the two cortical stages consists of a network of microcolumns (see §3.3.4.2) with

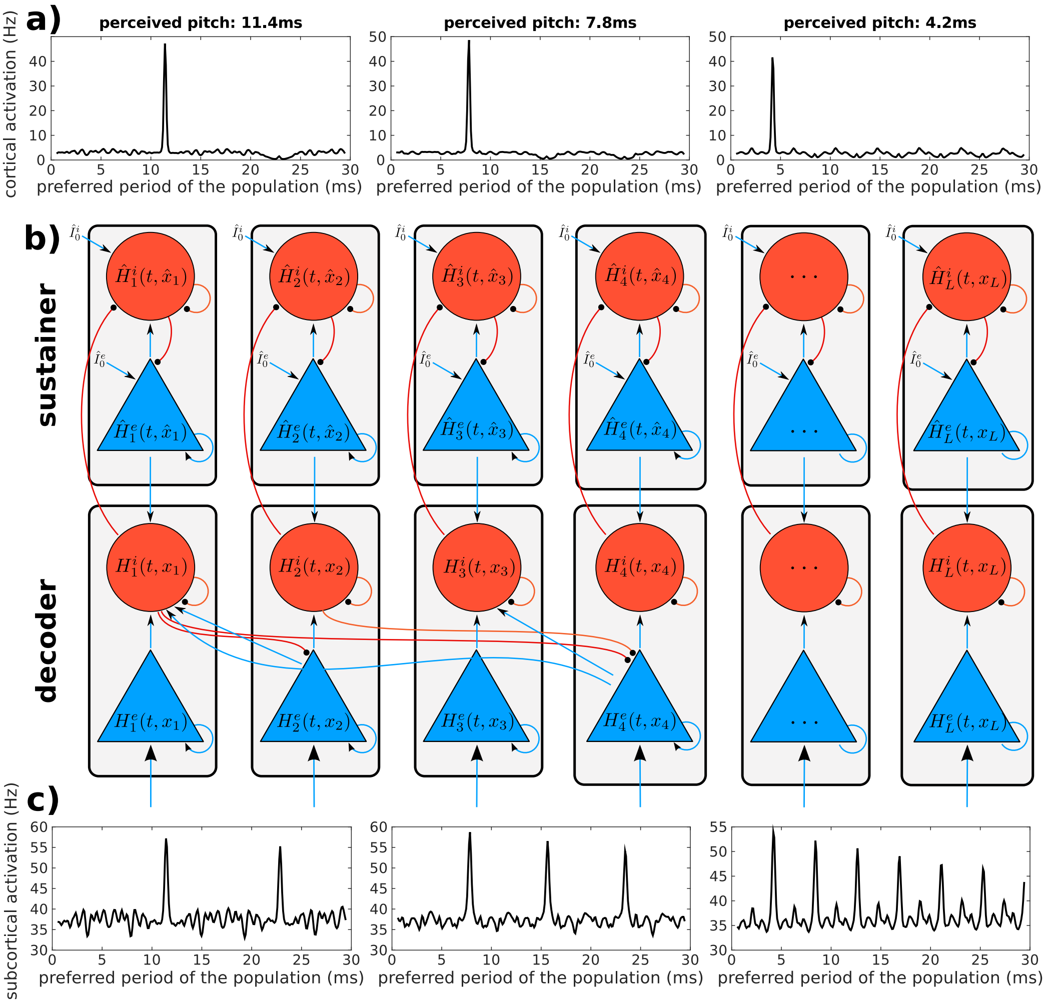


Figure 4.1: Basic schematics of the model. a) Cortical representation after the decoding of the autocorrelation patterns depicted in (c). b) Model diagram. The model consist of two networks, each with $N = 250$ columns (grey rectangles) modelled using one excitatory (triangles) and one inhibitory (circles) ensemble. Each block represents a given pitch value with periods ranging from 0.5 ms to 33 ms. The bottom network is the *decoder*, and the top network is the *sustainer*. Excitatory (e) and inhibitory i populations are characterised by their instantaneous average activity $H^{e,i}(t,x)$; hat notation is used to represent variables in the sustainer. Population activity depends on the total synaptic input of each population x; see details in §4.2.5.1. c) Average autocorrelation patterns for three IRNs with different pitch values.

preferred frequencies ranging from 0.5 ms (2000 Hz) [21,125] to 33 ms (30 Hz) [77]. Columns are modelled with two neural ensembles (see Figure 4.1): one excitatory, aggregating the properties of the spiny stellate and pyramidal excitatory populations [141] (see §3.3.4.2); and one inhibitory, comprising only inhibitory interneurons [141,150].

Together, decoder and sustainer effectively transform the subcortical input into a stable firing-rate representation, reminiscent of a receptive field [57,68,120] (see top plots in Figure 4.1A). The transformation is mediated by a specific structure of connectivity patterns of the cortical ensembles in the decoder layer (see Figure 4.2). Such a structure is designed to facilitate the inhibition of lower harmonics during the integration and is inspired by harmonic patterns of connectivity frequently reported in intracranial recordings of mammal auditory cortex (see [62] for a review).

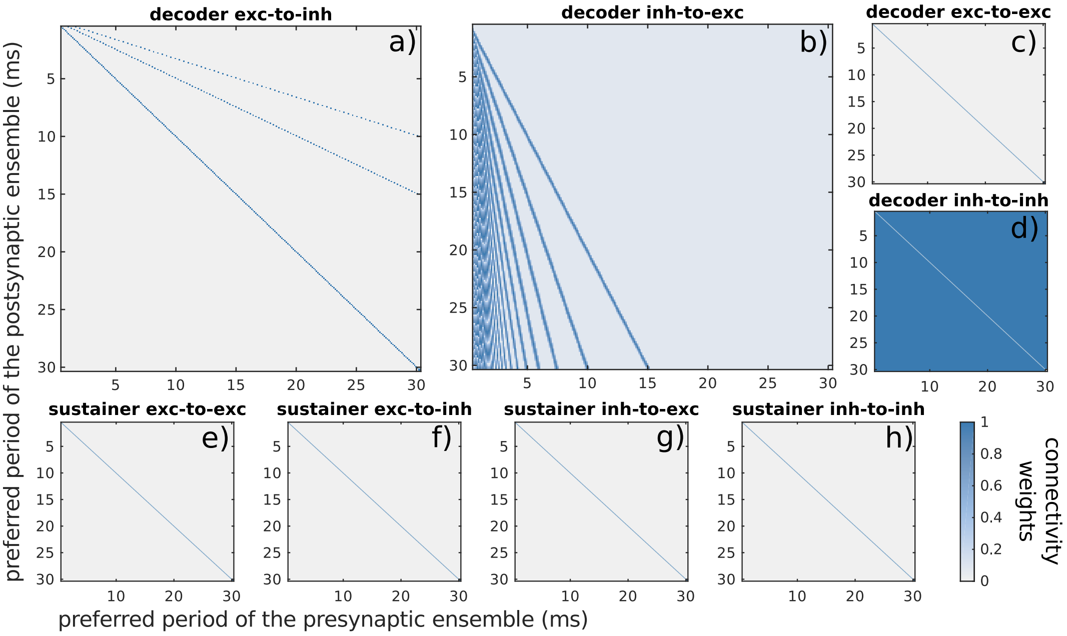


Figure 4.2: Connectivity weights. Matrices depict connectivity weights $C^{**}_{\alpha\beta} \in [0,1]$ between a presynaptic ensemble α (x-axis) and a postsynaptic ensemble β (y-axis). Matrices in the top correspond to connections C^{ei} (a), C^{ie} (b), C^{ee} (c), and C^{ii}(d), between ensembles in the decoder; bottom diagonal matrices, shown for completion, correspond to connections $\hat{C}^{ee}$ (e), $\hat{C}^{ei}$ (f), $\hat{C}^{ie}$ (g), and $\hat{C}^{ii}$ (h), in the sustainer. Superindices i and e are respectively used to denote inhibitory and excitatory ensembles.

The characteristic period of the column with the largest activity in the inhibitory ensembles in the decoder is predictive of the perceived pitch (but see §4.4.1). Pitch is as well coded in the inhibitory and excitatory activity at the sustainer, as will be further discussed. The equivalent dipole moment elicited in each of the two cortical networks is computed as the aggregated activity of the pyramidal excitatory cells [133].

The following sections describe the model in detail, explaining the functioning and predictive power of the different components of the processing system. Although some of the parameters of the model are specified in the text (specially for the subcortical system), values for the cortical parameters are all provided in Table 4.1 for the reader's convenience.

Peripheral and subcortical processing

Cortical input is generated according to the summary autocorrelation function [21,125] (see § 3.2.2.2 for more details), which is here assumed to yield the neural representation of pitch at subcortical stages.

Auditory nerve activity

Auditory nerve activity is generated using Zilany's model of the auditory periphery [39,117], a phenomenological model considering the non-linear response of the basilar membrane, the asymmetry of the cochlear channel bandwidths, and two distinct adaptation mechanisms (see Figure 3.1 and §3.1). Zilany's model presents a very efficient computational implementation that speeds up the simulations considerably; however, we do not expect significant differences at the cortical level when using similarly sophisticated peripheral models such as the Meddis comprehensive MAP [36] (see also §3.1).

Model parameters are set to consider 40 cochlear channels with centre frequencies between 125 Hz and 10 kHz, in agreement with the standard configuration used in previous autocorrelation models that are able to account for the elicited pitch of a wide range of stimuli [17].

Autocorrelation

Zilany's model returns the probability of spiking $p_k(t)$ at each instant t and cochlear channel k in response to a given stimulus. The SACF $A(t)$ associated to these spike trains is calculated following Equation (3.2):

$$\tau_n \dot{A}_n(t) = -A_n(t) + \sum_k p_k(t)\, p_k(t - \delta t_n)$$

where $n = 1 \dots N$ indexes the characteristic period δt_n of the autocorrelation output. Integration constants $\tau_n = 1.25\, \delta t_n$, with $\tau_n \leq 2.5\,\mathrm{ms}$ (see §3.3.1 and [125,131]).

The model considers $N = 250$ linearly distributed delays δt_n ranging from $\delta t_1 = 0.5\,\mathrm{ms}$, a conservative estimation of the phase-locking limit of the auditory nerve [64] (chosen here to avoid interference with the zero lag of the SACF [119]; see §3.2.2.3 for more details) to $\delta t_N = 33\,\mathrm{ms}$, near the lower limit of pitch $f_{\mathrm{min}} \sim 30\,\mathrm{Hz}$ [77].

As widely discussed in § 3.2.2.2 and §3.5.2, the SACF formulation yields prominent peaks of activation $A_n(t)$ at the delays characterising the harmonics of the period T characterising the stimulus' pitch; i.e, $\delta t_n = k\,T, \quad k = 0, 1, 2, \dots$. Classical patterns of activation are depicted in Figure 3.2.

Regularisation

Integration dynamics of the cortical model are sensitive to the absolute amplitude of the subcortical input: values under a certain threshold would fail to provoke a cortical reaction at all, whilst activity exceeding the range of operability triggers extreme responses that destabilise the model's response.

However, due to its lack of biological realisms, the SACF amplitude depends strongly on factors independent of pitch such as the spectral envelope and loudness of the stimuli, potentially inducing changes in the cortical dynamics that are not observed in the experimental data. The cortical model presented here focuses on the lemniscal ascending pathway, and hence elements of the auditory sensation such as loudness are not modelled. More importantly, the SACF is a highly idealized simplification of early representations of pitch and hence only provides an approximate input to auditory cortex. Thus, we suggest that only the relative height of the SAFC peaks, and not their absolute value, provides information about the input stimulus [124].

Thus, we introduced a regularisation procedure with two main targets. First, to remove the dependence of the absolute amplitude of the SACF peaks on the stimulus' loudness and timbre, guaranteeing that the cortical model response is only shaped by the relative differences (rather than the absolute value) of the SACF harmonic patterns of activation. Second, to reduce the variations induced by timbre in the signal-to-noise ratio of the SACF, also removing spurious activation as the one elicited by white noises without pitch. After this process, the regularised SACF is used as direct input for the cortical model.

SACF regularisation is carried out in three sequential steps: first, an adaptive multiplicative normalisation adjusts the overall heigh of the SACF peaks so that all SACFs show the same value at $\delta_t = 0$; second, a fixed additive term effectively subtracts the SACF baseline, attenuating signal-to-noise differences observed in SACFs elicited by sounds with different timbres; third, the corrected SACF is rescaled by a multiplicative factor that transforms the unit-less normalised SACF into units of firing rate (i.e, Hz). These three steps and their biological substrate are discussed below.

Active normalisation. Since the subcortical input is provided in an idealised fashion according to the principles of the SACF, the specific mechanisms underlying the normalisation of the thalamic input were not modelled. Instead, we took a phenomenological approach and divided the overall SACF response by a normalisation factor $Z(t)$ [148], chosen as the amplitude of the SACF peak at a hypothetical $\delta t_0 = 0$. The activity of $Z(t)$ is driven by the same dynamics as the SACF:

$$\tau_Z \dot{Z}(t) = -Z(t) + \sum_k p_k(t)^2 \tag{4.1}$$

where $\tau_Z = 2.5\,\text{ms}$ is the time constant corresponding to a zero lag [131]. Noisy stimuli like iterated rippled noises yield rapidly changing normalisation factors $Z(t)$ that affect the stability of the regularised SACF. To overcome this problem, both SACF and $Z(t)$ are further lowpass filtered by a leaky integrator [125].

The time constant of the lowpass filter was set to $\tau = 20\,\text{ms}$ in order to effectively smooth out phase-locked oscillations over 50 Hz that could have been induced in the cortical model through the SACF input. The frequency bound was set in agreement with electrophysiological results in humans reporting a decrease of phase-locking fidelity in auditory cortex over 50 Hz [16].

Baseline removal. Baseline removal is performed by subtracting a fixed $b_0 = 0.35$ of the normalised SACF:

$$A_n(t) \rightarrow \frac{A_n^{\text{low}}(t)}{Z^{\text{low}}(t)} - b_0$$

Baseline b_0 was chosen in such a way that a white noise elicits a negligible activation in the subcortical representation, according to fMRI results reporting pitch selective activation in cochlear nucleus and inferior colliculus [19, 50].

Rescaling. The normalised baseline-corrected SACF is then rescaled using a constant multiplicative factor $A_0/(1 - b_0)$, chosen such that the height of the zero-pole of the SACF (and thus the maximum possible value in the subcortical representation) equals A_0. According to a previous model of cortical perceptual integration [151], the rescaling parameter $A_0 = 75\,\text{Hz}$, yielding typical firing rates of $\sim 60\,\text{Hz}$.

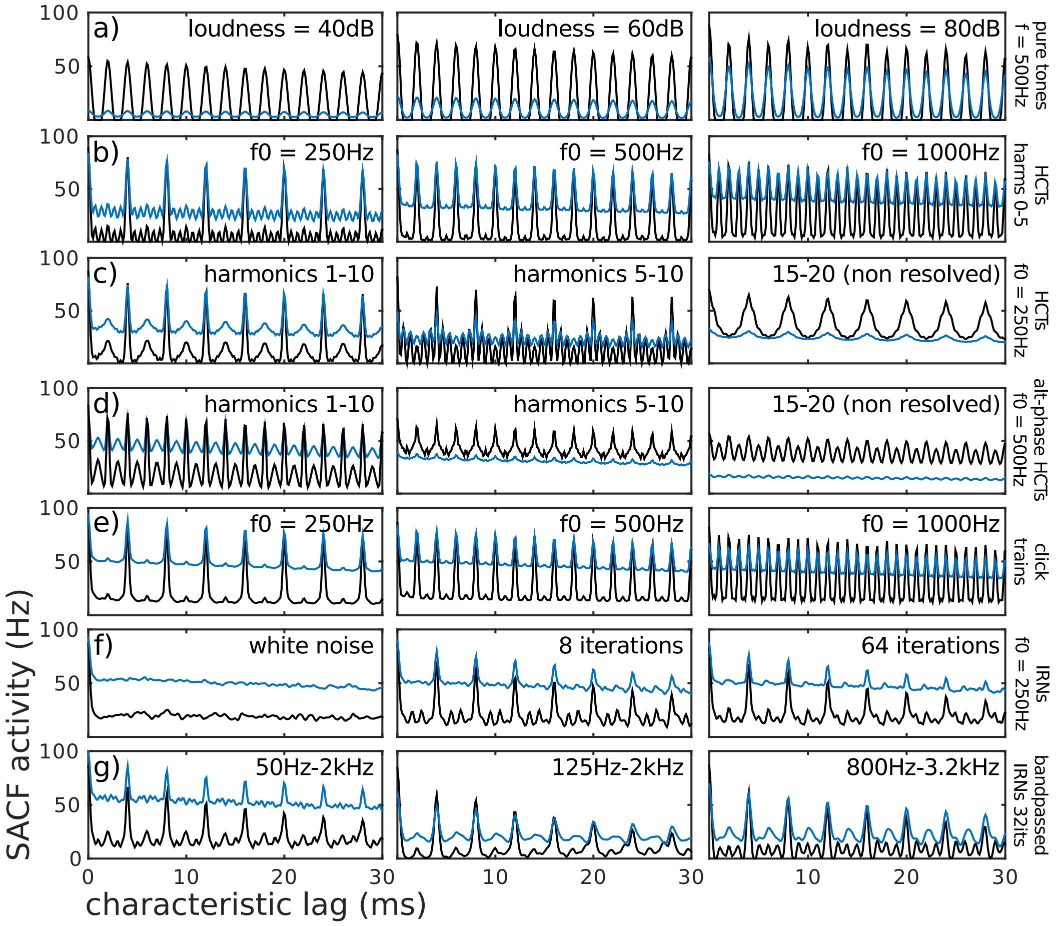


Figure 4.3: Effect of the regularisation procedure over the SACF. The figure compares several raw ($A_n(t)$, in blue) and regularised ($\hat{A}_n(t)$, in black) firing rates of the neurons hypothetically carrying the SACF representations. a) Pure tones ($f_0 = 500\,\mathrm{Hz}$) at different intensity levels. b) Harmonic complex tones with the first six harmonics and different fundamental frequencies. c) Harmonic complex tones eliciting virtual pitch (missing fundamental) with different spectral shapes; harmonics in the last panel are not independently resolved in the cochlea. d) Alternate-phase HCTs ($f_0 = 500\,\mathrm{Hz}$, see §2.4.1 and §3.2.2.5) with resolved and unresolved harmonics. e) Click trains with different fundamental frequencies. f) Iterated rippled noises with different number of iterations (delay $d = 4\,\mathrm{ms}$, equivalent to $f_0 = 250\,\mathrm{Hz}$); first panel shows a white noise (equivalent to 0 iterations). g) Bandpass filtered IRNs ($f_0 = 250\,\mathrm{Hz}$, 32 iterations) with different filtering configurations). When not specified, loudness was set to 80 dB (SPL). Non-regularised SACF was scaled with a factor 0.03 in all panels for visualisation purposes.

The regularised SACF $\hat{A}_n(t)$ is then defined as follows:

$$A_n(t) \to \hat{A}_n(t) = \frac{A_0}{1-b_0}\left(\frac{A_n^{\mathrm{low}}(t)}{Z^{\mathrm{low}}(t)} - b_0\right) \tag{4.2}$$

where $A_n^{\mathrm{low}}(t)$ and $Z^{\mathrm{low}}(t)$ are the lowpass filtered SACF and normalisation factor.

Biophysical biological substrate of the regularisation process. The operations involved in the regularisation process introduced above can be interpreted as instances of neuronal normalisation, a canonical operation underlying neural computation widely present in multiple neural systems [148]. Adaptive normalisation is typically exerted through global inhibition modulated by the overall activity in the network [148]. At a subcortical level, this inhibition could be driven by afferents from the nucleus reticularis thalami, a thin inhibitory layer covering the thalamus [152].

Baseline removal is a simple non-adaptive linear operation that could be accounted for by considering the input offset (I_0 in Equation 4.4 and Figure 4.5) of subcortical neurons integrating the ACF activity. The rescaling operation is used to transform the normalised values of the SACF into a quantity with a physical meaning, and it should be understood as a simple unit conversion.

The decoder

Decoder's architecture

The decoder consists of $N = 250$ interconnected cortical microcolumns, each one modelled as a circuit of two interacting neural ensembles: one excitatory (e), characterised by the average firing rate H_n^e of the excitatory neurons in the column n; and one inhibitory (i), characterised by H_n^i (see Figure 4.1B). Each excitatory ensemble n receives selective input from the corresponding subcortical channel $\hat{A}_n(t)$. A large activation in a column n is associated with a perceived pitch of δt_n [21].

Excitatory ensembles in the decoder do not connect with other excitatory ensembles in the network other than themselves, whilst inhibitory ensembles connect globally with other excitatory populations as shown in (Figure 4.1).

Connectivity is characterised by stronger inhibitory-to-excitatory connections from a population encoding the period δt_n with populations encoding any of the lower harmonics of such period (i.e, $k\,\delta t_n$, $k = 1, 2\ldots$; see full connectivity matrices in Figure 4.2). This connectivity architecture is inspired in harmonic connectivity patterns found in mammal auditory cortex [62, 153] and enables the system to facilitate the inhibition of the lower harmonics elicited during the peripheral processing.

Excitatory populations encoding δt_n connect to inhibitory populations encoding several higher harmonics $\delta t_n/k, \quad k = 1, 2, \ldots$ Inhibitory ensembles present uniform connections towards inhibitory ensembles in other blocks that shunt spurious and noisy inhibitory activity induced by their multiple excitatory inputs (see Figure 4.2).

Decoding process

The decoding process is summarised in Figure 4.4 for a stimulus with $f_0 = 250\,\mathrm{Hz}$, equivalent to $T = 4\,\mathrm{ms}$. First, the autocorrelation function extracts periodicities in the auditory nerve activity. A SACF channel becomes active ($\hat{A}_n(t) > 0$) after $t = 1.25\,\delta t_n$. Thus, the first peak of activation arises in the SACF only after $t = 5\,\mathrm{ms}$ (see Figure 4.4A) and propagates to the decoder layer, eliciting activation in the excitatory population H_n^e, characterised by $\delta t_n = 4\,\mathrm{ms}$ (see Figure 4.4B). 5 ms later, the SACF presents another peak at n', with $\delta t_{n'} = 8\,\mathrm{ms}$, which in turn activates the excitatory population $H_{n'}^e$ (see Figures 4.4A and B).

After the third peak appears at $A_{n''}$ and propagates to the decoder, the overall input from populations H^e_n, $H^e_{n'}$, and $H^e_{n''}$ towards their common inhibitory target H^i_n is large enough to elicit a strong activation of such inhibitory ensemble (see Figure 4.4C). In turn, H^i_n strongly inhibits the excitatory ensembles of characteristic frequencies corresponding to its lower harmonics. Thus, the increased activation of H^i_n results in the inhibition of excitatory populations at n', n'', and successive low harmonics (see Figure 4.4B), effectively transforming the harmonic patterns of the SACF into a single-peak representation.

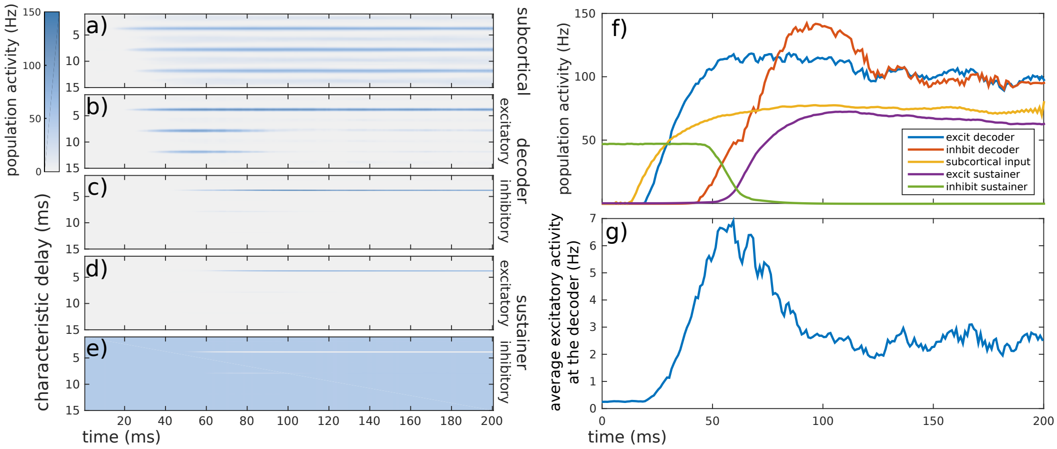


Figure 4.4: Illustration of the decoding process during the processing of an IRN. Plots show the evolution of the key variables of the model during the processing of the first 200 ms of an IRN eliciting a pitch $f_0 = 250\,\text{Hz}$, equivalent to a period $T = 4\,\text{ms}$. a)–e) Show the evolution of cortical and subcortical populations encoding characteristic delays between 0.5 ms and 15 ms. a) Shows the activity of subcortical populations $\hat{A}^n(t)$. b) and c) show the excitatory $H^e_n(t)$ and inhibitory $H^i_n(t)$ activities at the decoder, respectively. d) and e) show the excitatory $\hat{H}^e_n(t)$ and inhibitory $\hat{H}^i_n(t)$ activities at the sustainer. f) Shows the activity evolution at the ensembles encoding the pitch value, characterised by a delay $d = 4\,\text{ms}$. g) Shows the aggregated activity across all excitatory ensembles of the decoder; this quantity is monotonically related to the elicited auditory fields (see §4.2.7.1).

Decoding onset

The decoding process is designed to progressively build up evidence of SACF-like harmonic patterns characterising different pitch values. Requiring typically three peaks of the harmonic series is a reasonable balance between efficiency (the resolution of each additional peak takes around one extra period) and robustness (for instance, requiring only two peaks would yield spurious activations in common higher harmonics in dyads, as will be shown in Chapter 5). Moreover, resolving three peaks in the subcortical system takes around four periods of the stimulus (the integration time constant of a subcortical population encoding a delay δt_n is $1.25\,\delta t_n$ [131]), which explains why tone's durations of four times the period are necessary for robust pitch discrimination [43] and why the pitch onset response latency scales with four times the tone's period [22].

The sustainer

The functional role of the sustainer

As explained above, the decoding process is based on the inhibition of the activity at the lower harmonics present in the subcortical representation, $n', n'', \ldots$. Inhibition towards those peaks is driven by the inhibitory population H_n^i, which in turn is fed-forward by the excitatory inputs n' and n'' (see Figure 4.1B); i.e, precisely the ones that are being inhibited. Thus, right after the decoding process is triggered, these excitatory ensembles are no longer active and H_n^i quickly loses its driving input, causing the inhibition to decrease back to baseline levels. Without the action of the inhibition, the harmonic peaks propagating from the subcortical input would rise again in the cortical representation, ultimately triggering a new decoding process. This happens even when the subcortical input does not present any obvious discontinuity.

If cortical pitch processing would rely on a decoder of this kind, pitch would be extracted from the subcortical representation over and over during the tone's duration, eliciting a rather discontinued sensation. However, pitch of continuous tones is experience as a continuous sensation [43], and the auditory evoked fields show a single pitch onset response in cortex [105, 107].

The role of the *sustainer* network is to maintain a stable pitch representation in the decoder once it has been extracted from the subcortical representation. The sustainer reinforces the input at the inhibitory population n characterising the pitch, by replacing the input from the inhibited excitatory populations at the lower harmonics n' and n'' through an inter-layer recurrent process described below.

Sustainer's architecture

Like the decoder, the sustainer consists of $N = 250$ microcolumns modelled as a circuit with an excitatory ensemble, characterised by its average firing rate $\hat{H}_n^e$, and a inhibitory ensemble, characterised by $\hat{H}_n^i$ (see Figure 4.1B). Unlike the decoder, sustainer columns do not communicate with each other: ensembles only connect to ensembles in their same column (see Figure 4.2)

Ensembles at the sustainer receive direct input from their counterparts at the decoder: excitatory populations at the decoder H_n^e connect with excitatory populations at the sustainer $\hat{H}_n^e$, and inhibitory populations at the decoder H_n^i connect with inhibitory populations at the sustainer $\hat{H}_n^i$ (see Figure 4.1B). Moreover, sustainer's inhibitory and excitatory ensembles receive a constant background input form other cortical areas I_0^{sus}.

Sustaining process

In the absence of external input, the sustainer network rests at equilibrium with a steady activation in the inhibitory populations and no excitatory activity (see Figure 4.4D and E). Then, combined excitatory and inhibitory input from a given column n in the decoder decreases the activity in the inhibitory ensemble $\hat{H}_n^i$ of the sustainer (see Figure 4.4E) and enhances the activity of the excitatory population $\hat{H}_n^e$ at the sustainer layer (Figure 4.4D).

Top-down efferents connect each excitatory population $\hat{H}_n^e$ at the sustainer to its inhibitory counterpart at the decoder layer H_n^i. Thus, the activity propagated to the sustainer effectively results in a net top-down input towards the inhibitory population H_n^i, that replaces the input formerly received from the shunted lower harmonics n' and n'' (note the slight increase in the inhibitory activity in Figure 4.4F right after the rise of the sustainer's excitatory activity, caused by the short overlap between the decoder's and sustainer's inputs, and how the inhibitory ensemble is kept active even after the higher harmonics have been completely shunted from the decoder representation).

In summary, the sustainer maintains the already elicited inhibition at the decoder, effectively holding a continuous representation of the decoded pitch, in agreement with physiological recordings and perceptual observations.

Dynamic equations

Rate dynamics of the neural populations

Ensembles are modelled according to a mean-field approximation, which assumes that neurons at a given population have similar neural properties and they all receive the same input [30, 154]. Under this assumption, the average spiking rate $H(t)$ of the neurons in an ensemble evolves as the spiking rate of an average neuron whose parameters are the average of the parameters of the neurons across the population [30, 154].

We used a neural rate model derived from a leaky integrate-and-fire (LIF) neuronal model [154]. The model describes the evolution of the average firing rate $H(t)$ as a leaky integration of a transfer function $\phi(I)$ that maps the total input current to the response firing rate of the neuron at equilibrium. The transfer functions used here were derived empirically by Wong and Wang [151, 155] using simulations of a spiking network of LIF excitatory and inhibitory neurons.

The temporal evolution of the firing rates $H_n^e(t)$ (excitatory) and $H_n^i(t)$ (inhibitory) follows the dynamics of a canonical leaky integrator:

$$\tau^{\text{pop}}\,\dot{H}_n^{e,i}(t) = -H_n^{e,i}(t) + \phi^{e,i}(I_n^{e,i}(t)) \tag{4.3}$$

with transfer functions $\phi^{e,i}(I_n^{e,i}(t))$ [151] (see also Figure 4.5):

$$\phi^{e,i}(I) = \frac{a^{e,i}I - b^{e,i}}{1 - e^{-d^{e,i}(a^{e,i}I - b^{e,i})}} \tag{4.4}$$

Parameters of the excitatory and inhibitory transfer functions (a^e, b^e and d^e for the excitatory; a^i, b^i and d^i for the inhibitory) were taken from the original study by Wong and Wang [151]. The total synaptic inputs $I_n^e(t)$ and $I_n^i(t)$ are defined bellow in §4.2.5.3.

Dynamics of excitatory and inhibitory ensembles at the decoder and sustainer followed the same formulation. Equations above are thus valid for both the decoder ($H_n^{e,i}(t)$) and the sustained ($\hat{H}_n^{e,i}(t)$) populations, using the appropriate synaptic input ($I_n^{e,i}(t)$ for the decoder, $\hat{I}_n^{e,i}(t)$ for the sustainer) in each case.

Dynamics were numerically simulated using Euler's method with a time-step of 0.1 ms.

Adaptive time constants. Classical rate models are based on filters that approximate the population response to a given input [154, 156], but fail to capture how the neural activity shapes the dynamics of the response to the input. Ostojic and Brunel [156] developed an adaptive rate model based on the exponential integrate and fire (EIF) neural model [157, 158], where the filter is shaped according to the ensemble's activity through an adaptation of the effective integration time constant τ^{pop}:

$$\tau^{\text{pop}}(H(t)) = \tau_0^{\text{pop}}\,\Delta_T\,\frac{\phi'(I(t))}{H(t)} \tag{4.5}$$

where Δ_T is the sharpness of the action potential initiation in the EIF model and $\phi'(I(t))$ is the slope of the transfer function (see Equation (4.4)) at the current synaptic input $I(t)$.

Although our population dynamics are based on a slightly different neural model, we use their derivation as an approximation, based on the general observation that populations of neurons are more sensitive to input variations when they present a large firing rate [30, 156]. Their derivation is easily transferable to our model by considering that time constants are equivalent in rate models and membrane potential models [159], and using a small $\Delta_T \ll 1 = 0.05\,\mathrm{mV}$ reflecting that, unlike the EIF, the LIF model approximates the action action potential initiation as instantaneous [157].

Synapse dynamics

Neural communication is mediated through synapses. Synaptic gates open at the arrival of an action potential from the pre-synaptic neuron, releasing neurotransmitters that increase or decrease the membrane potential in the post-synaptic neurons. Here, we consider three kinds of neurotransmitters widely linked to cortical processes and perceptual integration [151, 155, 160, 161]: two excitatory, NMDA and AMPA; and one inhibitory, GABA. We model synaptic dynamics according to Brunel and Wang's classical derivation [160].

Synaptic gates driven by AMPA $S_n^{\mathrm{AMPA}}(t)$ and GABA $S_n^{\mathrm{GABA}}(t)$ neurotransmitters present fast dynamics and are modelled as leaky integrators with instantaneous rising times [160] and different decay times $\tau_{\mathrm{AMPA}} = 2\,\mathrm{ms}$ and $\tau_{\mathrm{GABA}} = 5\,\mathrm{ms}$ [160]. Synaptic gates are triggered by the activity in excitatory and inhibitory populations, respectively:

$$\begin{aligned} \dot{S}_n^{\mathrm{AMPA}}(t) &= -\frac{S_n^{\mathrm{AMPA}}(t)}{\tau_{\mathrm{AMPA}}} + H_n^e(t) + \sigma\nu_n^{\mathrm{AMPA}}(t) && (4.6)\\ \dot{S}_n^{\mathrm{GABA}}(t) &= -\frac{S_n^{\mathrm{GABA}}(t)}{\tau_{\mathrm{GABA}}} + H_n^i(t) + \sigma\nu_n^{\mathrm{GABA}}(t) && (4.7) \end{aligned}$$

Additive noise is introduced in the system as a Gaussian process $\nu_n(t)$ independently sampled for each synapse and instant t. Noise weight was set to $\sigma = 0.0007\,\mathrm{nA}$, according to the specifications of the original population model [151].

NMDA-driven synapses present slow dynamics and a finite rising time [160]:

$$\dot{S}_n^{\mathrm{NMDA}}(t) = -\frac{S_n^{\mathrm{NMDA}}(t)}{\tau_{\mathrm{NMDA}}} + \gamma\,(1 - S_{\mathrm{NMDA}}(t))\,H_n^e(t) + \sigma\nu_n(t) \qquad (4.8)$$

NMDA time constant was set to $\tau_{\mathrm{NMDA}} = 30\,\mathrm{ms}$ and the coupling parameter $\gamma = 0.641$ was taken from the literature [151, 160].

As in the previous section, equations driving the gating dynamics are the same in the decoder ($S_n^{\mathrm{NMDA}}(t)$, $S_n^{\mathrm{AMPA}}(t)$, and $S_n^{\mathrm{GABA}}(t)$, with $H_n^{e,i}(t)$) and the sustainer ($\hat{S}_n^{\mathrm{NMDA}}(t)$, $\hat{S}_n^{\mathrm{AMPA}}(t)$, and $\hat{S}_n^{\mathrm{GABA}}(t)$, with $\hat{H}_n^{e,i}(t)$).

Synaptic inputs

Total synaptic inputs to populations at the decoder $I_n^{i,e}(t)$ and the sustainer $\hat{I}_n^{i,e}(t)$ convey all the synaptic drive of the neurons, here divided for convenience in three separate contributions: internal input I_{int}, comprising inputs from populations within the same network; external input I_{ext}, exerted by sources from other networks; and a constant input drive I_0:

$$\begin{aligned} I_n^{i,e}(t) &= I_{n,\mathrm{int}}^{i,e}(t) + I_{n,\mathrm{ext}}^{i,e}(t) + I_{n,0}^{i,e}(t) && (4.9)\\ \hat{I}_n^{i,e}(t) &= \hat{I}_{n,\mathrm{int}}^{i,e}(t) + \hat{I}_{n,\mathrm{ext}}^{i,e}(t) + \hat{I}_{n,0}^{i,e}(t) && (4.10) \end{aligned}$$

Internal inputs. Internal inputs are defined as the sum of all synaptic outputs from the populations placed in the same network as the postsynaptic ensemble. Weights of the synaptic conductivity between two ensembles are encoded in the connectivity matrices $C^{ee}, C^{ei}, C^{ie}, C^{ii}$ in the decoder network and $\hat{C}^{ee}, \hat{C}^{ei}, \hat{C}^{ie}, \hat{C}^{ii}$ in the sustainer network. Connectivity weights $C_{\alpha\beta}^{**} \in [0, 1]$ are plotted in Figure 4.2 (as above, stars are used as wildcards for excitatory e or inhibitory i). Using these definitions, the following equations describe the explicit internal inputs $I_{\mathrm{int}}(t)$:

$$\begin{aligned} I^e_{n,\mathrm{int}}(t) &= \sum_k C^{ee}_{nk} \left(J^{ee}_{\mathrm{NMDA}}\, S^{\mathrm{NMDA}}_k(t) + J^{ee}_{\mathrm{AMPA}}\, S^{\mathrm{AMPA}}_k(t)\right) \\ &\quad - \sum_k C^{ie}_{nk} J^{ie}_{\mathrm{GABA}}\, S^{\mathrm{GABA}}_k(t) \end{aligned} \tag{4.11}$$

$$\begin{aligned} I^i_{n,\mathrm{int}}(t) &= \sum_k C^{ei}_{nk} \left(J^{ie}_{\mathrm{NMDA}}\, S^{\mathrm{NMDA}}_k(t) + J^{ei}_{\mathrm{AMPA}}\, S^{\mathrm{AMPA}}_k(t)\right) \\ &\quad - \sum_k C^{ii}_{nk} J^{ii}_{\mathrm{GABA}}\, S^{\mathrm{GABA}}_k(t) \end{aligned} \tag{4.12}$$

$$\begin{aligned} \hat{I}^e_{n,\mathrm{int}}(t) &= \sum_k \hat{C}^{ee}_{nk} \left(\hat{J}^{ee}_{\mathrm{NMDA}}\, \hat{S}^{\mathrm{NMDA}}_k(t) + \hat{J}^{ee}_{\mathrm{AMPA}}\, \hat{S}^{\mathrm{AMPA}}_k(t)\right) \\ &\quad - \sum_k C^{ie}_{nk} \hat{J}^{ie}_{\mathrm{GABA}}\, \hat{S}^{\mathrm{GABA}}_k(t) \end{aligned} \tag{4.13}$$

$$\begin{aligned} \hat{I}^i_{n,\mathrm{int}}(t) &= \sum_k \hat{C}^{ei}_{nk} \left(\hat{J}^{ei}_{\mathrm{NMDA}}\, \hat{S}^{\mathrm{NMDA}}_k(t) + \hat{J}^{ei}_{\mathrm{AMPA}}\, \hat{S}^{\mathrm{AMPA}}_k(t)\right) \\ &\quad - \sum_k C^{ii}_{nk} \hat{J}^{ii}_{\mathrm{GABA}}\, \hat{S}^{\mathrm{GABA}}_k(t) \end{aligned} \tag{4.14}$$

Conductivities J^{**}_{NMDA}, J^{**}_{AMPA}, J^{**}_{GABA}, $\hat{J}^{**}_{\mathrm{NMDA}}$, $\hat{J}^{**}_{\mathrm{AMPA}}$, and $\hat{J}^{**}_{\mathrm{GABA}}$, were initialised to typical values in the literature $J \simeq 0.15\,\mathrm{nA}$ [151] and fine-tuned to ensure the model displayed the desired dynamics (parameter fitting is thoroughly described in Appendix A).

External inputs. External input received by excitatory ensembles at the decoder consists of the thalamic input provided by the regularised SACF (as described in §4.2.2.3) implemented subcortically. Thalamic input is transmitted to cortex by means of AMPA-driven synapses, according to previous studies in perceptual integration [151]:

$$I^e_{n,\mathrm{ext}}(t) = J^{th}_{\mathrm{AMPA}}\, S^{th,\mathrm{AMPA}}_n(t) \tag{4.15}$$

The conductivity J^{th}_{AMPA} was adjusted to ensure a smooth propagation of the subcortical input to the cortical populations (see §A). The thalamic AMPA gating variables $S^{th,\mathrm{AMPA}}_n(t)$ followed the dynamics described in Equation (4.6), using the firing rate of the regularised SACF output $\hat{A}_n(t)$ as the driver for the AMPA release:

$$\dot{S}^{th,\mathrm{AMPA}}_n(t) = -\frac{S^{th,\mathrm{AMPA}}_n(t)}{\tau_{\mathrm{AMPA}}} + A_n(t) \tag{4.16}$$

Inhibitory ensembles at the decoder receive external input from the top-down efferents coming from the sustainer. Top-down excitatory processes in cortex are often linked to NMDA dynamics [162], so GABAergic synapses were not considered here:

$$I^i_{n,\mathrm{ext}}(t) = J^e_{\mathrm{NMDA}}\, \hat{S}^{th,\mathrm{NMDA}}_n(t) \tag{4.17}$$

The efferent conductivity J^e_{NMDA} was adjusted to make the top-down reinforcement of the inhibitory ensembles at the decoder strong enough to replace the excitatory input from the shunted lower harmonics after the decoding process (once again, see §A).

Sustainer's external inputs are bottom-up afferents sourced in the decoder network, driven by GABAergic (inhibitory) and AMPAergic (excitatory) synapses [151, 162]:

$$\hat{I}^{e}_{n,\mathrm{ext}}(t) \;=\; \hat{J}^{a}_{\mathrm{AMPA}}\, S^{\mathrm{AMPA}}_{n}(t) \tag{4.18}$$

$$\hat{I}^{i}_{n,\mathrm{ext}}(t) \;=\; \hat{J}^{a}_{\mathrm{GABA}}\, S^{\mathrm{GABA}}_{n}(t) \tag{4.19}$$

Afferent conductivities $\hat{J}^{a}_{\mathrm{AMPA,\ GABA}}$ were set to make the sustainer both sensitive to decoded decisions and robust to spurious activations (see §A).

Constant input drive. Constant inputs in the decoder $I^{e}_{n,0}(t) = I^{e}_{0}$ and $I^{i}_{n,0}(t) = I^{i}_{0}$ were chosen to make the system reactive to the subcortical input, without eliciting spontaneous activity in the network (see Figure 4.5).

An additional constant drive $I^{\mathrm{sus}}_{0} = 0.24\,\mathrm{nA}$ was applied to the populations at the sustainer (see §4.2.4.2): $\hat{I}^{e,i}_{n,0}(t) = I^{e,i}_{0} + I^{\mathrm{sus}}_{0}$.

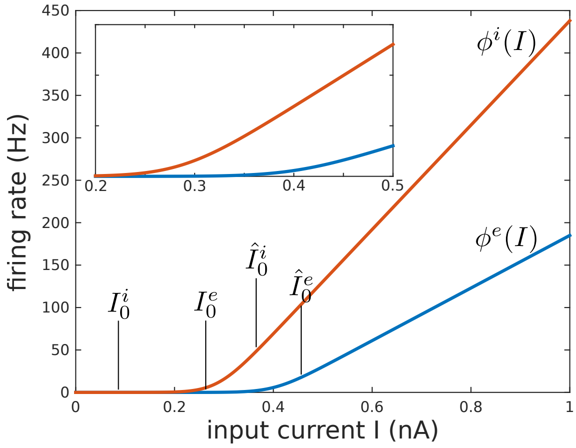


Figure 4.5: Transfer functions $\phi(I)$ and constant input currents I_0. The miniature in the top is a detail of the non-linear rise of the transfer functions. Analytical formulation of the transfer functions is shown in Equation 4.4.

Adaptation

All cortical ensembles present an adaptation term taking into account short-term habituation, effectively regulating the maximum firing rate a population can reach. Adaptation is modelled in a phenomenological fashion, as a negative input current added to the synaptic drive of each ensemble $I_n(t)$. Adaptation $I^{\mathrm{adap}_n(t)}$ effective currents evolve as a canonical leaky integrator fed by the population's activity [154]:

$$\dot{I}^{\mathrm{adap}(t)}_{n} = -\frac{I^{\mathrm{adap}(t)}_{n}}{\tau^{\mathrm{adap}}} + \alpha H_n(t) \tag{4.20}$$

with adaptation time constant $\tau^{\mathrm{adap}} = 100\,\mathrm{ms}$ [154] and adaptation strength $\alpha = 3 \times 10^{-6}$. Same equations and parameters drive the adaptation dynamics of excitatory and inhibitory ensembles at the sustainer and the decoder.

Connectivity matrices

Connectivity matrices were designed to facilitate the detection of SACF-like patterns of activation and the further inhibition of the higher harmonics, but they present a few tunable parameters. The formulation of the matrices is shown bellow, actual values of the model are provided in Figure 4.2. Harmonic connectivity patterns were inspired by findings in intracranial recordings, reporting strong connections between frequency selective neurons whose frequencies were harmonically related (see [62] for a review).

Connectivity weights at the decoder Excitatory ensembles at the decoder layer carry the subcortical representation of the stimulus. In order to keep this representation as faithful as possible, excitatory populations do not excite other excitatory ensembles within the decoding network other than themselves; thus, the *ee* connectivities at the decoder are only recurrent and can be expressed as a Kronecker delta:

$$C^{ee}_{\alpha\beta} = \delta_{\alpha\beta} \tag{4.21}$$

Excitatory ensembles at the decoder excite inhibitory ensembles characterising higher harmonics (i.e, at columns n', n'', ... characterising periods $\delta t_{n'} = \delta t_n/2$, $\delta t_{n''} = \delta t_n/3, \ldots$; see §4.2.3.1 and §4.2.3.2):

$$C^{ei}_{\alpha\beta} = \begin{cases} 1 & \text{if } \frac{\delta t_\alpha}{\delta t_\beta} = k, \quad k = 1, 2, \ldots, K^{ei} \\ 0 & \text{otherwise.} \end{cases} \tag{4.22}$$

where $K^{ei} - 1 = 2$ is the number of higher harmonics each excitatory ensemble targets.

Inhibitory ensembles shunt excitatory populations encoding lower harmonics (i.e, at columns n', n'', ... characterising periods $\delta t_{n'} = 2\delta t_n$, $\delta t_{n''} = 3\delta t_n, \ldots$):

$$C^{ie}_{\alpha\beta} = \begin{cases} 1 & \text{if } \frac{\delta t_\alpha}{\delta t_\beta} = k, \quad k = 2, 3, \ldots, K^{ie} \\ c^{ie}_0 & \text{otherwise.} \end{cases} \tag{4.23}$$

where $K^{ie} = 66$ is the number of lower harmonics each inhibitory ensemble targets. The number was chosen so that the population encoding the shortest period $\delta t_1 = 0.5\,\text{ms}$ would inhibit all its lower harmonics up to the longest considered period $\delta t_N = 33\,\text{ms}$. Excitatory ensembles not encoding harmonics of the presynaptic inhibitory ensemble also receive a subtle inhibition $c^{ie}_0 \ll 1$ that is taken as a free parameter of the model (c^{ie}_0 tuning criteria are provided in §A).

Inhibitory to inhibitory recurrent weights are chosen as to avoid self-inhibition:

$$C^{ii}_{\alpha\beta} = 1 - \delta_{\alpha\beta} \tag{4.24}$$

Connectivity weights at the sustainer Sustainer's connections were essentially local: ensembles communicate only with ensembles in their same column. Thus,

$$\hat{C}^{ee}_{\alpha\beta} = \hat{C}^{ei}_{\alpha\beta} = \hat{C}^{ie}_{\alpha\beta} = \hat{C}^{ii}_{\alpha\beta} = \delta_{\alpha\beta} \tag{4.25}$$

Derivation of the evoked fields

Elicited equivalent dipole moments

Assuming that all microcolumns within each of the two cortical networks present similar orientations, the total dipolar moment representing the neuromagnetic field elicited by each network is proportional to the aggregated excitatory activity along the network [133, 134]. In the decoder:

$$m(t) = \sum_n H^e_n(t + \Delta t_{\text{subcort}}) \tag{4.26}$$

The subcortical delay $\Delta t_{\text{subcort}}$ accounts for the time elapsed from tone onset until the signal first arrives in primary auditory cortex and was fixed to $\Delta t_{\text{subcort}} = 70\,\text{ms}$ using the latency of the POR elicited by an IRN with a delay of 8 ms as reference (see §4.3.2.1). To account for trial to trial variability, we further averaged the predicted dipole moment across several runs $M(t) = \langle m(t) \rangle_{\text{runs}}$. Dipole moments at the sustainer were represented by $\hat{m}(t)$ and $\hat{M}(t)$.

Decoder elicited field

The aggregated response at the decoder shows a large transient, peaking around $\sim 100\,\text{ms}$, after which the field stabilises to a low activity level. The build up of the peak is a consequence of the different harmonics of the SACF propagating into the decoder network. After enough information (i.e, a sufficient number of peaks) is available to decode the pitch value from the SACF representation, inhibitory ensembles begin to effectively shunt the lower harmonics. The transient peaks at the instant where inhibition overcomes the subcortical input; the equilibrium state is achieved when all the lower harmonics have been inhibited (see Figure 4.4 and the associated generated field in the subpanel F). We identified these onset dynamics with the pitch onset response in anterolateral Heschl's gyrus (see §2.3.2.4). A comparison between the decoder's derived field and the observed equivalent magnetic dipole elicited in alHG by an IRN is shown in Figure 4.6.

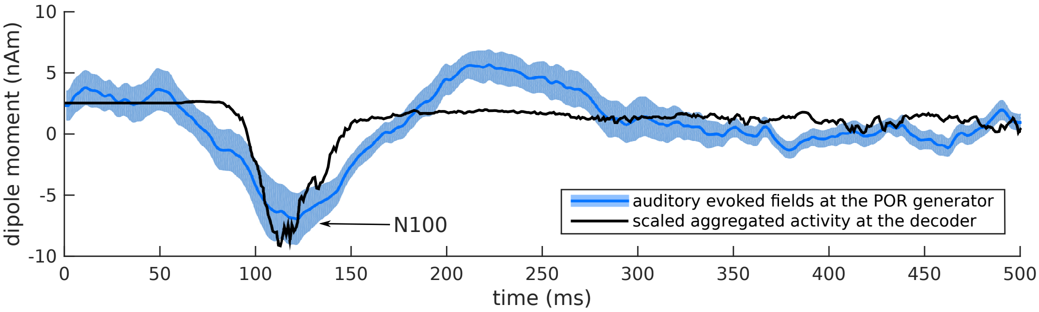


Figure 4.6: Comparison of the model's simulated fields and the POR dynamics in an MEG recording for an IRN.. Stimulus consisted of an IRN with 32 iterations and a delay $d = 4\,\text{ms}$. Data corresponds to the equivalent dipole moment $m(t)$ (see Equation (4.26)) elicited in the POR generator as a response to an IRN eliciting the same pitch; the equivalent fields were scaled by a negative linear factor accounting for the monotonic relationship between $m(t)$ and the elicited fields that was fitted for this particular example.

Sustainer elicited field

The build up at the sustainer begins much later, right after the inhibitory onset in the decoding network, and holds steady until its decay, which starts a few milliseconds after stimulus' offset. Due to this late onset and the observed offset delay, we identified the neuromagnetic field elicited by the sustainer with the pitch-related sustained field of the auditory evoked fields (see examples in Figures 4.16 and 4.17), whose generator is typically found adjacent to the POR generator [107, 108] (see also §2.3.2.6).

Parameter selection

After fixing the structure of the connectivity weight matrices and the normalisation parameters of the subcortical input, the dynamics of the cortical model still depend on 36 different parameters. 17 of those parameters were fixed according to the literature; the remaining parameteres were adjusted using a five-stage procedure described in Appendix A. Table 4.1 lists the final parameter values and the stage where they were tuned; tuning criteria are provided in the appendix.

Results

Psychophysics

The model's perceptual predictions were evaluated for a wide range of stimuli encompassing pure tones, harmonic complex tones, click trains, and iterated rippled noises. Stimuli were generated using our own scripts, which are included in the model's libraries. Sample rate was set to 100 kHz; onset and offset were smoothed using a $\tau = 2.5\,\mathrm{ms}$ Hamming window [5].

Model's perceptual output is provided by the temporal average of the inhibitory activity in the most salient column in the decoder layer after the decoding process (e.g. Figure 4.7). Perceived pitch is robustly encoded in the inhibitory ensembles of the decoder layer, and hence in the excitatory and inhibitory populations in the sustainer layer, as discussed above. Thus, the model's readout can be indistinctly set to any of these three populations. Inhibitory neurons at the decoder show a faster response; excitatory neurons at the sustainer, although responding a few milliseconds later, show excitatory activity that can be easily transmitted to higher centres in the auditory hierarchy.

An average was taken from $t_0 = 150\,\mathrm{ms}$ to $t_1 = 250\,\mathrm{ms}$ to make explicit the robustness of the extracted pitch representation, although plotting the raw model's responses at any time after the end of the decoding process yields similar results. Results for the populations of the sustainer are precisely correlated to the activity of the inhibitory populations at the decoder and are thus generally not reported to avoid redundancy.

Stimuli modalities were tested for 31 different pitch values ranging from 66 Hz to 1614 Hz in a piecewise asymptotic scale (i.e, piecewise linear in the space of the periods). The highest tested pitch frequency corresponds to the characteristic frequency of the last functional population of the model. The lower tested frequency was chosen 35 Hz, well below the limit of optimal operability of the model corresponding to $\sim 100\,\mathrm{Hz}$ (i.e. the minimum frequency at which three harmonics $\delta t = 10\,\mathrm{ms}$, $\delta t' = 20\,\mathrm{ms}$ and $\delta t'' = 30\,\mathrm{ms}$ are present in the SACF representation) to test its behaviour in suboptimal conditions.

Pure tones

Figure 4.7 shows the response at subcortical and cortical populations of the model averaged between $t_0 = 150\,\mathrm{ms}$ and $t_1 = 250\,\mathrm{ms}$. Each row of the heat maps shows the response in different columns to a certain stimulus. Activity in both, the decoder's and sustainer's ensembles, shows a unimodal distribution for each stimulus centred on the populations that encode the characteristic period corresponding to the period of the sinusoid. Perceptual predictions are particularly robust for the inhibitory populations of the decoder and excitatory and inhibitory populations of the sustainer, which are in full agreement with classical perceptual results in pure tones [5] for frequencies over $\sim 125\,\mathrm{Hz}$. The lack of responses under $\sim 125\,\mathrm{Hz}$ is due to the lower frequency limit of the peripheral model [163].

As discussed earlier (see §4.2.4.3), responses in the sustainer are precisely correlated to the responses of the decoder's inhibitory ensembles. However, excitatory populations at the decoder still preserve aspects of the spectral representation of the regularised SACF. This is partly due to the loss of precision introduced by the discretisation of the period space: when decoding the pitch value T of the stimulus, inhibitory ensembles characterising the period δt closest to the actual stimulus period T become active, introducing a small error

parameter	value	fittin stage / source
J^{th}_{AMPA}	2.5 nA	stage 1
J^{e}_{NMDA}	0.45 nA	stage 4
$\hat{J}^{a}_{\mathrm{GABA}}$	0.45 nA	stage 3
$\hat{J}^{a}_{\mathrm{AMPA}}$	0.35 nA	stage 4
J^{ee}_{NMDA}	0.14 nA	stage 1
J^{ee}_{AMPA}	9.9×10^{-4} nA	[151]
J^{ei}_{NMDA}	0.19 nA	stage 2
J^{ei}_{AMPA}	6.5×10^{-5} nA	[151]
J^{ie}_{GABA}	0.66 nA	stage 2
J^{ii}_{GABA}	0.11 nA	stage 2
$\hat{J}^{ee}_{\mathrm{NMDA}}$	0.25 nA	stage 3
$\hat{J}^{ee}_{\mathrm{AMPA}}$	9.9×10^{-4} nA	[151]
$\hat{J}^{ei}_{\mathrm{NMDA}}$	0.00 nA	stage 3
$\hat{J}^{ei}_{\mathrm{AMPA}}$	9.9×10^{-4} nA	[151]
$\hat{J}^{ie}_{\mathrm{GABA}}$	0.80 nA	stage 3
$\hat{J}^{ii}_{\mathrm{GABA}}$	0.00 nA	stage 3
c^{ie}_{0}	0.1	stage 4
γ	0.641	[160]
a^{e}	310 $(\mathrm{VnC})^{-1}$	[151]
b^{e}	125 Hz	[151]
d^{e}	0.16 s	[151]
a^{i}	615 $(\mathrm{VnC})^{-1}$	[151]
b^{i}	177 Hz	[151]
d^{i}	0.087 s	[151]
I^{e}_{0}	0.315 nA	stage 1
I^{i}_{0}	0.15 nA	stage 2
$\hat{I}^{e}_{0}$	0.26 nA	stage 3
$\hat{I}^{i}_{0}$	0.18 nA	stage 3
τ_{AMPA}	2 ms	[160]
τ_{GABA}	5 ms	[160]
τ_{NMDA}	30 ms	ad-hoc
τ^{pop}	10 ms	[156]
Δ_{T}	0.05 mV	ad-hoc
σ	0.007 nA	[151]
τ^{adap}	100 ms	[154]
α	3×10^{-6} nA	stage 1

Table 4.1: Values for the parameters used in the cortical model. Last column specifies whether if the parameter was fitted (and, in that case, at which step of the fitting process was it fixed; see Appendix A) or if the value was selected ad-hoc (and, if the value was taken from the literature, the specific source that was used).

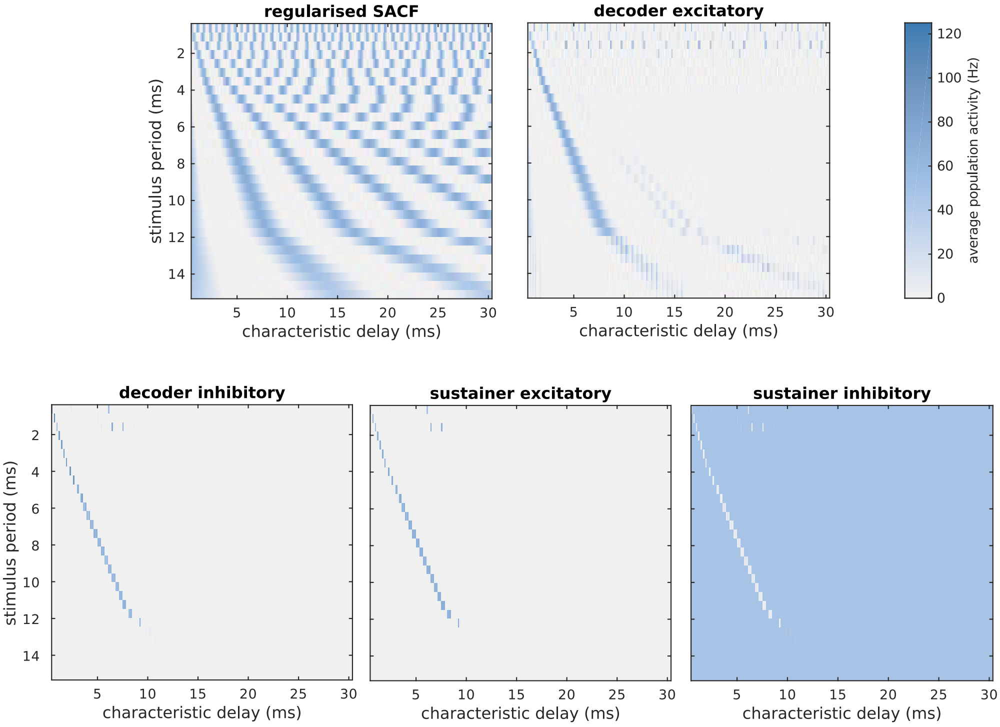


Figure 4.7: Perceptual predictions for pure tones. Heat maps represent different stages of the model (subcortical and decoder's excitatory ensembles at the top, decoder's inhibitory and sustainer's ensembles at the bottom). Each row shows the average activity $\langle H(t)\rangle^{t\in(150,250)\,\mathrm{ms}}$ elicited by each given stimulus. The piece-wise linear pattern observed for the maximum activation peaks is to the equally piece-wise linear distribution of pitch values chosen for the experimentation.

in the decoded pitch $\Delta = \|\delta t - T\|$. The most active inhibitory ensemble represents the decoded pitch with a relatively high precision. However, this error escalates linearly when considering the inhibitory connections towards the excitatory ensembles characterising the lower harmonics of T; specifically, the inhibition of the nth harmonic shows a deviation $\Delta_n = \|n\delta t - nT\| = n\Delta$. This effect is specially prominent in tones with high pitch values (see top rows in the decoder excitatory ensembles of the heat map in Figure 4.7). A similar effect is observed for low pitched tones (Figure 4.7) due to the outstanding thickness of the harmonic peaks elicited by pure tones in the SACF representation. Note that the discretisation error would not be present in an actual system shown a continuous pitch representation.

Despite these effects, the model provides a robust representation of the decoded pitch values at all times after the short transient of the decoding process.

Harmonic complex tones, click trains, and iterated rippled noises

Figure 4.8 shows the perceptual output of the model to harmonic complex tones with and without missing fundamentals, and both with resolved and unresolved harmonics. Results show consistent pitch predictions, fully in line with classic perceptual results [5]. Like in the results corresponding to pure tones, decoder excitatory ensembles show some characteristics of the SACF representations, partly due to the discretisation error described above. A new effect is shown in the excitatory activity elicited by unresolved harmonics in the decoder layer (see Figure 4.8C), where frequencies not corresponding to the harmonic series, and thus not inhibited by the decoding mechanism, are present in the cortical representation (see the parallel lines surrounding the main harmonic peaks in the SACF).

Note that, since in this case higher harmonics are represented in the cochlea, the model does decode pitch values under the lower limit of the peripheral system.

Figure 4.9 shows the perceptual output of the model to iterated rippled noises with different configurations. Results show consistent pitch predictions, fully in line with classic perceptual results (e.g, [5, 11]). Bandpass filtering removes short-time correlations that are crucial to characterise IRNs with short delays, resulting in the observed degradation of the SACF representations for the higher pitched IRNs (see Figures 4.9B and 4.9C).

Alternating-phase harmonic complex tones

Next, we tried to predict the perceived pitch evoked by alternating-phase harmonic complex tones with missing fundamentals. These stimuli elicit a pitch sensation equivalent to their fundamental frequency when its harmonics are independently resolved in the cochlea, but a sensation equivalent to twice their fundamental frequency when its harmonics are not independently resolved (see §2.2.3.2 and §3.2.2.5).

Perceptual results for ALT-HCTs were similar than those displayed in Figure 4.8, and did not show significant differences when their harmonics were resolved or unresolved in the cochlea, failing to reproduce perceptual observations reported in the literature for the unresolved condition [5].

We identified the cause of this problem in the large baseline value (i.e, activity common to all delays) observed in the regularised SACF (see Figure 4.10, left panels). A possible solution to this caveat is to adjust the SACF baseline parameter b_0 in order to increase the relative salience of the peaks (see justification in §4.2.2.3).

Figure 4.10 illustrates the effect of the baseline adjustment; the figure shows the corresponding perceptual results when running the model using $b_0 \to \hat{b}_0 = 2\,b_0$. Here, complexes comprising harmonics 11 to 14 present a clear shift from the fundamental frequency to the first higher harmonic for frequencies over 200 Hz. This shift is due to the presence of the secondary peaks in the regularised SACF (see rows corresponding to stimuli with periods 1–5 ms in Figure 4.10). Under the 200 Hz limit, the smaller peaks in the SACF characterising the higher harmonics lie below the increased baseline limit set before and the model prediction fails, predicting instead an elicited pitch corresponding to the fundamental frequency of the HTCs (see the broken line on the top left corner of the rightmost panels in Figure 4.10B).

Thus, a dynamic adjustment of the baseline removal parameter b_0 is necessary in order to fully predict the pitch elicited by alternating phase HCTs.

Shifted HCTs

Lastly, we computed the perceptual predictions of the model to shifted HCTs, a set of stimuli consisting of a harmonic complex $(f_0, 2f_0, 3f_0, \dots)$ in which the frequency of each harmonic is shifted by a fixed amount Δ: $(f_0 + \Delta, 2f_0 + \Delta, 3f_0 + \Delta, \dots)$. Paradoxically, shifted HCTs elicit a pitch percept close to their fundamental frequency f_0 with a pitch deviation that is much smaller than the shifted quantity Δ [164].

In order to test the performance of the model on these stimuli we computed 20 shifted HCTs with the same fundamental frequency $f_0 = 100$ Hz and 20 linearly distributed shifts Δ ranging from 0 to 100 Hz. Results are shown in Figure 4.11.

The model showed a good agreement with the perceptual data for small deviations $\Delta \sim 0$–30 Hz and $\Delta \sim 70$–100 Hz. Stimuli with shifts $\Delta \sim 30$–70 Hz result in an octave shift

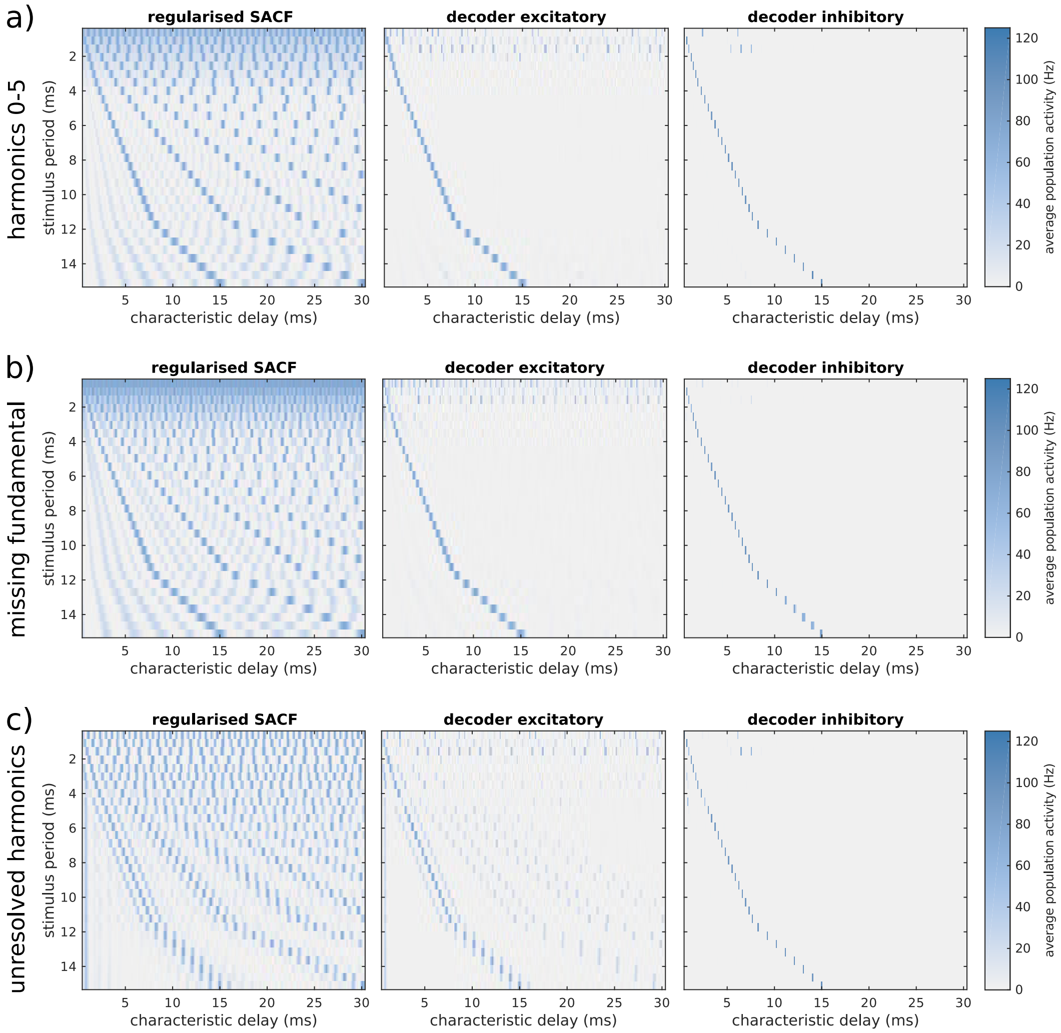


Figure 4.8: Perceptual predictions for harmonic complex tones. Heat maps represent different stages/ensembles of the model (sustainer's responses, precisely correlated with the decoder's inhibitory responses, were omitted for simplicity). a) HCTs with the fundamental frequency f_0 and subsequent five harmonics. b) HCTs with a missing fundamental, consisting on harmonics f_1 to f_5. c) HCTs with a missing fundamental and harmonics not independently resolved in the cochlea, generated as an HCT with 50 harmonics further bandpass filtered between 3.2 kHz and 5 kHz. Note that, since the activity at the decoder inhibitory populations is equivalent for HCTs with and without the fundamental, the readout of the model cannot be use to perform judgements on pitch salience (see also §6.2.3.2).

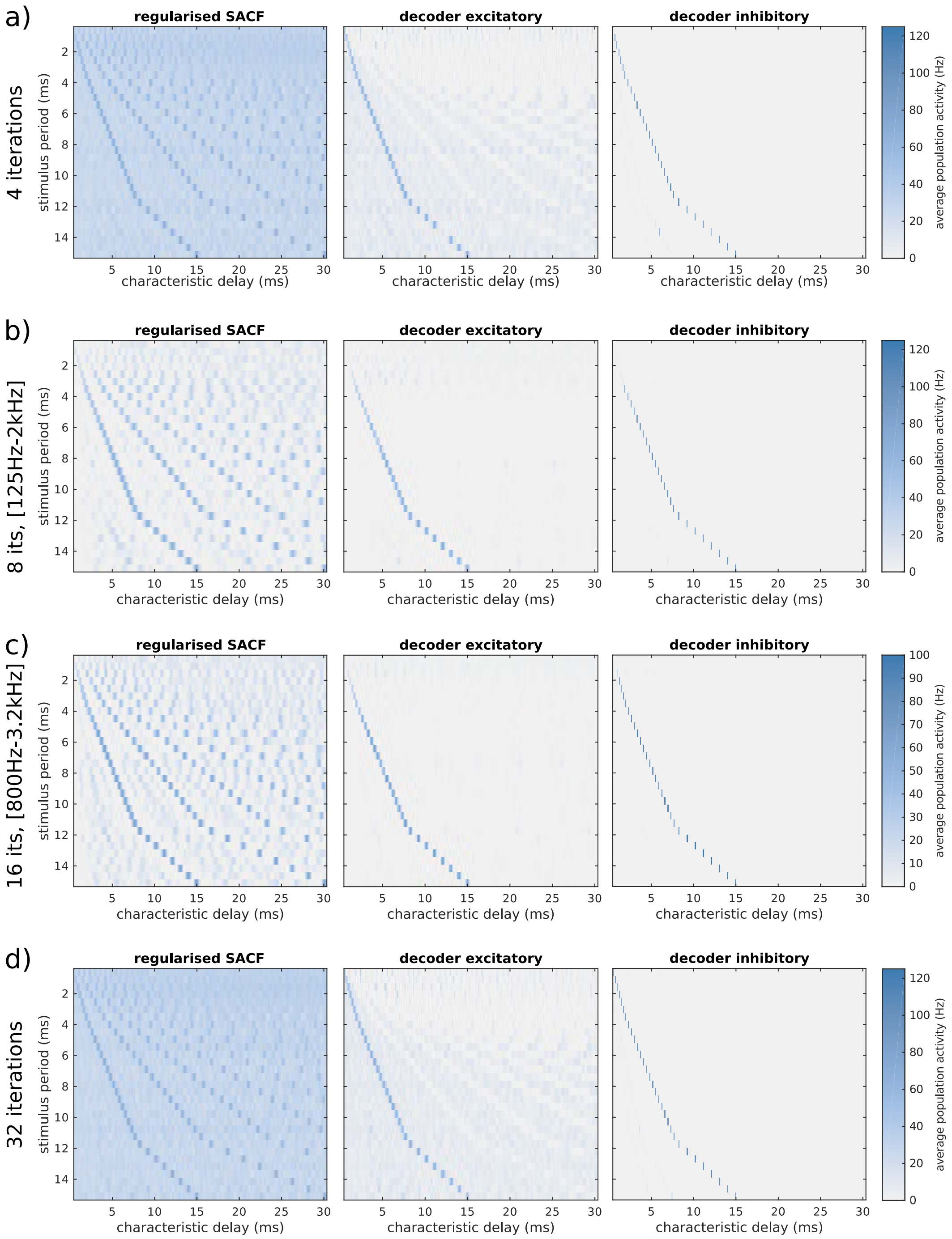


Figure 4.9: Perceptual predictions for iterated rippled noises. a) IRNs with 4 iterations. b) IRNs with 8 iterations, bandpass filtered between 125 Hz and 2 kHz; configuration was chosen according to the specifications of IRN dyads used in §5.2. c) IRNs with 16 iterations, bandpass filtered between 0.8 kHz and 3.2 kHz; filter parameters were chosen according to [22]. D) IRNs with 32 iterations.

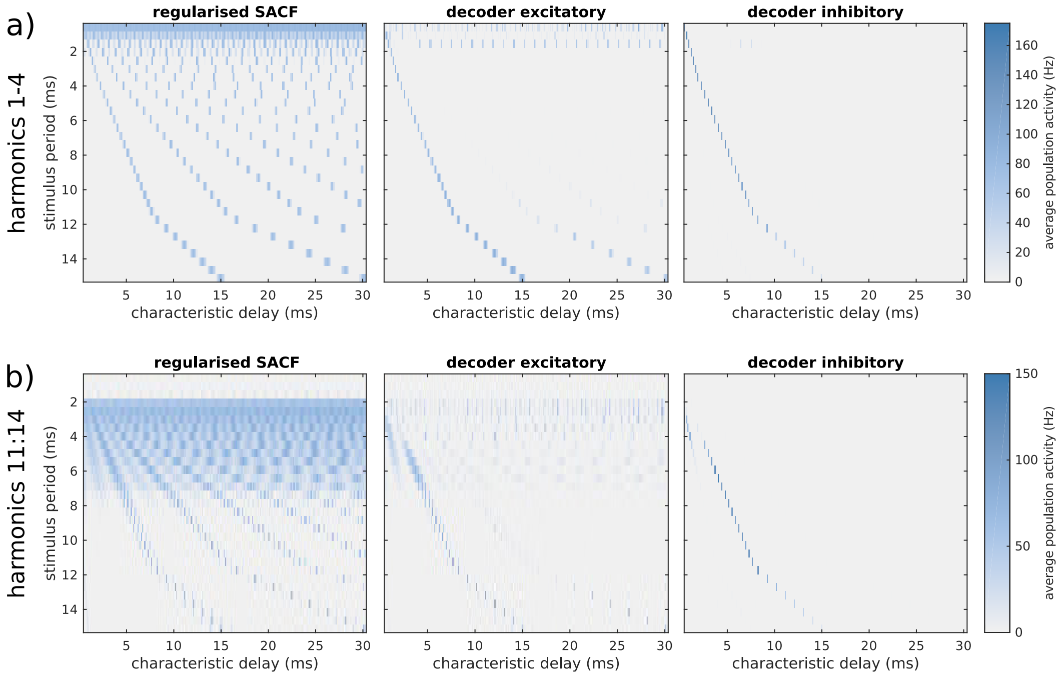


Figure 4.10: Perceptual predictions for alternated-phase HCTs. a) Complexes comprising harmonics f_1 to f_4. b) Complexes comprising harmonics eleven to fourteen, out of the range of cochlear resolvability. Simulations were performed using an alternative baseline for the SACF $\hat{b}_0 = 0.70$ (see §4.2.2.3).

(note that the ratios in that range are close to 1/3 and 1/2 of the fundamental frequency) that is not reported in the perceptual data [164]; however, it should be noted that available data corresponds to a single subject study that might have overlook an otherwise apparent octave shift.

POR dynamics

POR morphology dependence with pitch in IRNs

The association of the decoder's aggregated excitatory activity with the dipole moment at the generator of the pitch onset response in alHG (see §4.2.7.2) allows to perform qualitative predictions on the POR morphology.

Figure 4.12A shows the IRN latency predictions of our model for a series of IRNs with 16 iterations bandpassed between 0.8 kHz and 3 kHz, in comparison with the observed latency values of the POR evoked by the same stimuli [22]. Despite the large variability shown by the model predictions due to cortical noise and subcortical trial-to-trial variability, the prediction values are fully in line with the experimental observations.

The observed POR latency correlation with pitch is a consequence of the dependence of the build up of inhibitory activity in the decoder network with the stimulus fundamental period. Inhibition of the population encoding the period of the perceived pitch is typically triggered after the third peak in the harmonic series is represented in cortex (see §4.2.3.3).

Next, we computed the model's latency predictions for five IRNs with different number of iterations but eliciting the same pitch value $T = 16\,ms$. Comparison between the simulations and experimental data are shown in Figure 4.12B. IRN latency predictions show that the

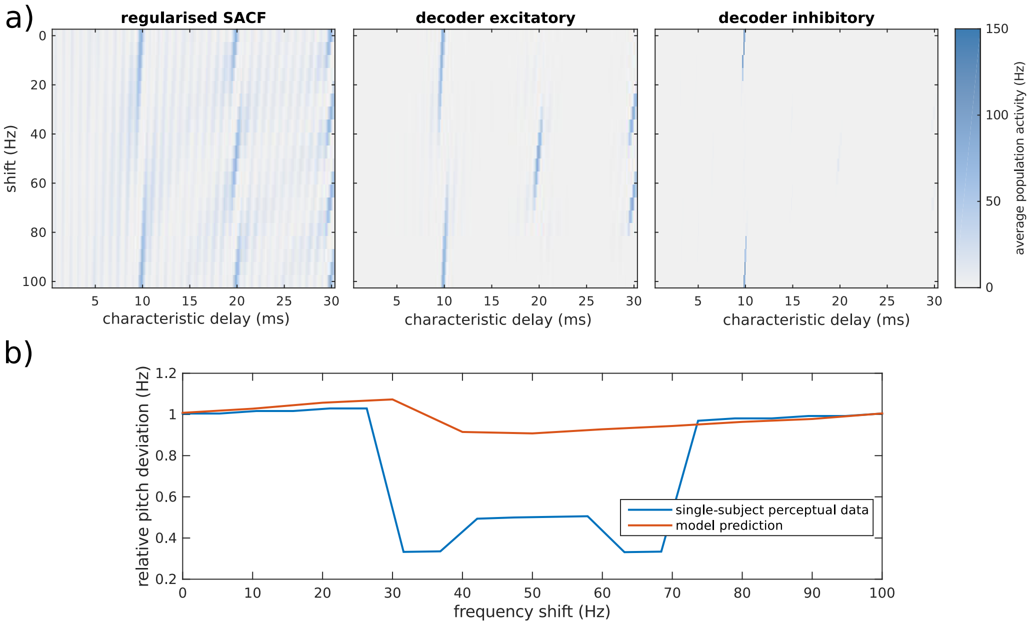


Figure 4.11: Perceptual predictions for shifted HCTs. a) Heat maps show the model responses for each of the applied shifts. b) Ratio predicted/perceived pitch and the pitch corresponding to the fundamental $f_0 = 100$. Predicted pitch was computed as the characteristic period of the decoder's inhibitory population with the largest activation. Perceived pitch corresponds to the elicited pitch reported by a single subject for the same set of stimuli; perceptual data from [164], Fig 1.

pitch strength of the IRNs does not significantly affect the POR's latency, in full agreement with experimental observations. The not-significant but noticeable trend observed in the simulations is a reflection of the slight increase in processing time provoked by the decrease in the signal-to-noise ratio in the SACF representations associated with a low number of iterations.

POR morphology dependence with the number of iterations in IRNs

The number of iterations of an IRN is correlated with its elicited pitch strength; both quantities are reflected in the POR's amplitude [22]. IRNs with a with a large number of iterations evoke larger transients than IRNs with a low number of iterations. Moreover, when fixing the number of iterations, the PORs amplitude does not seem to be affected by its elicited pitch [22]. Experimental data is shown in black in Figure 4.14.

Although pitch strength is out of the scope of our study, it is worth to test the capacity of the model to perform predictions of the POR amplitude. IRNs with a large number of iterations result in robust harmonic SACF representations, where each peak of the series presents a similar height. In contrast, IRNs with a low number of iterations elicit a degraded harmonic series with a prominent first peak and lower activity levels in the subsequent peaks of their SACF representation (see Figure 3.2C). Since the POR amplitude depends on the aggregated activity at the second and third peaks of the SACF, the differences in the subcortical representation might explain the POR amplitude dependence with this parameter.

The aggregated excitatory activity at the decoder is monotonically related to the equivalent dipole moment that would have been elicited by a neural implementation of the decoder's

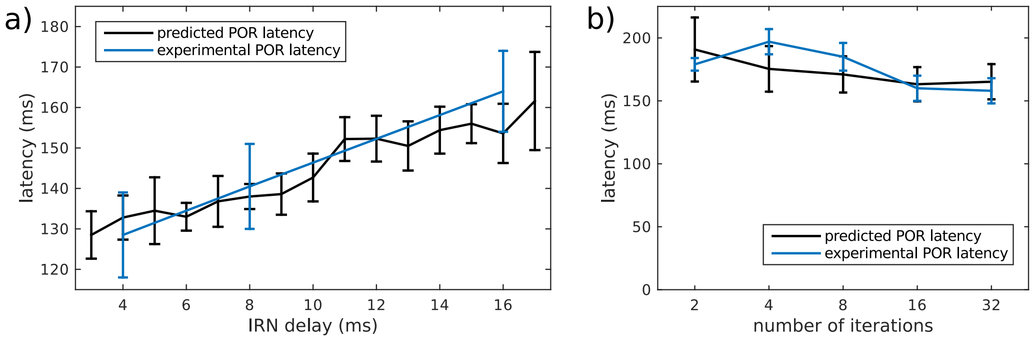


Figure 4.12: POR morphology dependence on pitch value in IRNs. a) POR latency predictions for IRNs eliciting different pitch values in comparison with experimental observations. b) POR latency predictions for IRNS with different number of iterations in comparison with experimental observations. Experimental data was taken from [22] (Figure 4). Predictions were averaged across 5 runs of the model; error bars are standard errors. Experimental data was taken from [22] (Figure 4).

network in cortex (see S4.2.7.1); although the exact relationship between one and the other is unknown, qualitative predictions of the POR's amplitude can be performed on this basis. Thus, we compared the overall activity in the decoder's excitatory populations at the simulated POR's peak elicited by a series of IRNs with the POR's amplitude elicited by the same stimuli; see Figure 4.14. In agreement with the available experimental data, the simulated field was only significantly affected by the IRNs number of iterations (see Figure 4.14B), but not by their elicited pitch (see Figure 4.14A).

POR and N100 latency in pure tones

Next, we tested the model capacity to predict the POR latency of pure tones. Latency predictions are shown in Figure 4.15A, in comparison with experimental data for the same set of stimuli [91].

Since the POR and EOR elicited by stimuli other than IRN cannot be experimentally disentangled from each other, the validation of our latency predictions for such stimuli is challenging: our model predicts the POR behaviour whilst MEG studies report the latency of the N100 [91]. We addressed this problem by expressing the N100 latency as an average between the POR's and the EOR's respective latencies. This simple model allowed us to correct our POR predictions by assuming that both transient had similar weights in the average (e.g. they both have the same amplitude and their generators are equally distant from the N100 dipole) and an EOR latency of 95 ms. Corrected values are shown in Figure 4.15B, in full agreement with experimental observations.

Sustained field dynamics

To conclude, Figures 4.16 and 4.17 show the average activity at the sustainer during the processing of several stimuli eliciting different pitch sensations. Each figure compares the aggregated dynamics of the sustainer with the aggregated activity at the decoder (which provided the feedforward input to the sustainer) and the aggregated cortical input (which provided the input to the decoder). Comparison with experimental data is challenging because current MEG techniques do not allow for a complete separation of the adjacent dipoles of the POR and the sustained field.

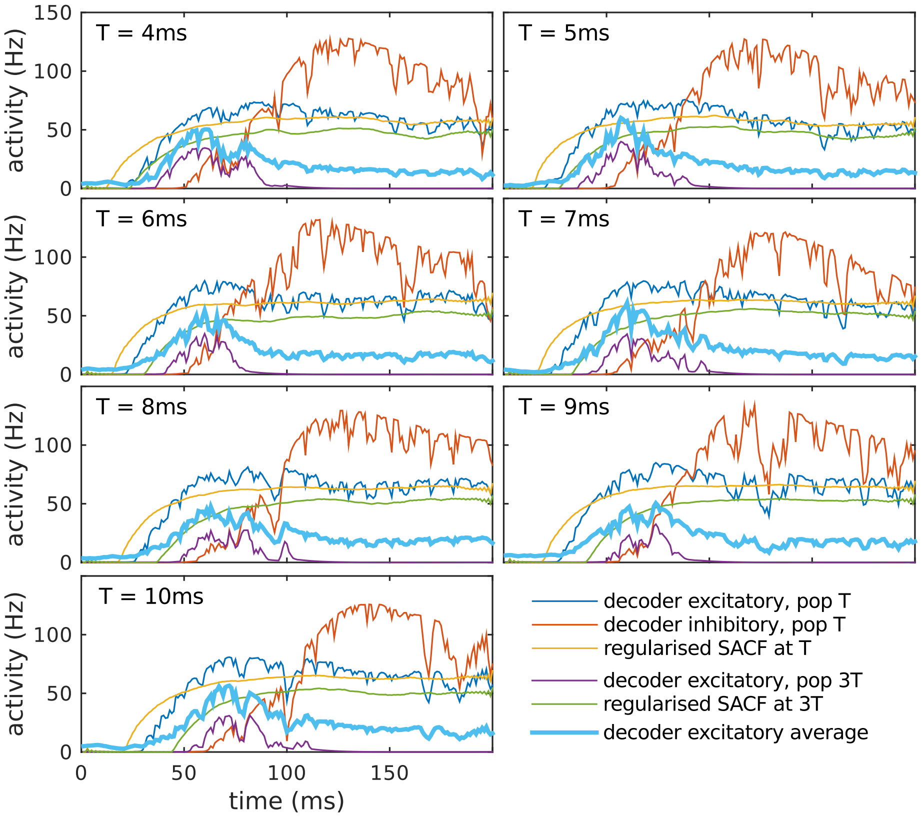


Figure 4.13: Evolution of the system during pitch processing for several IRNs. Plots show an example of the explicit evolution of the activity of several populations during the decoding of the pitch of IRN stimuli eliciting different pitch sensations T. Dark blue, red, and yellow lines show the evolution of the decoder's excitatory, decoder's inhibitory, and subcortical populations encoding the pitch of the stimulus $\delta t_l = T$; purple and green show the evolution of the decoder's excitatory, and subcortical populations encoding the third peak of the harmonic series $\delta t_l = 3T$; clear blue shows the averaged excitatory activity in the decoder (i.e, the predicted elicited field associated to the decoder network). Stimuli were generated using the same parameters as in Figure 4.12.

Whilst the decoder's response to an increase of activity in the subcortical input is almost immediate (see steep increase in Figure 4.16), the sustainer's onset is only observed after the arrival of POR's peak; i.e, shortly after the perceptual decision has been performed at the decoder, around 50–100 ms after the rise of the decoder's field, depending on the pitch of the stimulus.

The sustainer's field shows as well a significant delay in comparison with the offset of the cortical input. This delay is provoked by the inertia of the recurrent inhibitory-to-inhibitory connections between the decoder and the sustainer. The delay is approximately constant and roughly equal to $\delta \sim 50$ ms. This could quantitatively explain the offset delay observed by Gutschalk and colleagues [107] in the pitch-related sustained field (see §2.3.2.6). However, our simulations show a linear dependence with the stimulus' period of the subcortical offset with a factor one, whilst Gutschalk observations seem to set the dependence of the SF's offset delay on twice the period of the stimulus.

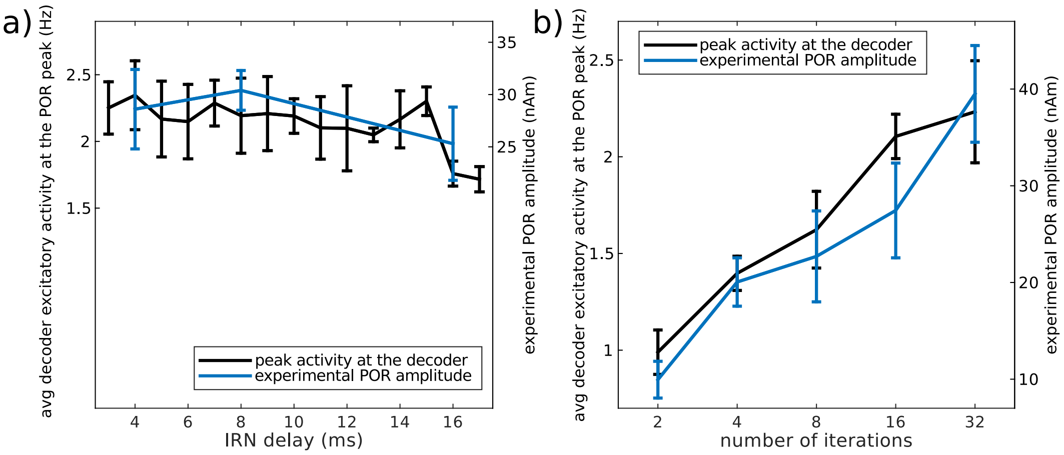


Figure 4.14: PORs amplitude dependence on the number of iterations of IRNs. Decoder's aggregated excitatory activity at the POR's peak in comparison with experimental observations on the POR's amplitude for a series of IRNs. a) IRNs with 16 iterations eliciting different pitch values. b) IRNS with different number of iterations and same delay $d = 16$ ms, eliciting the same pitch percept. Experimental data was taken from [22] (Figure 4). Predictions were computed across 5 runs of the model; error bars are standard errors. Stimuli were the same as in Figure 4.12.

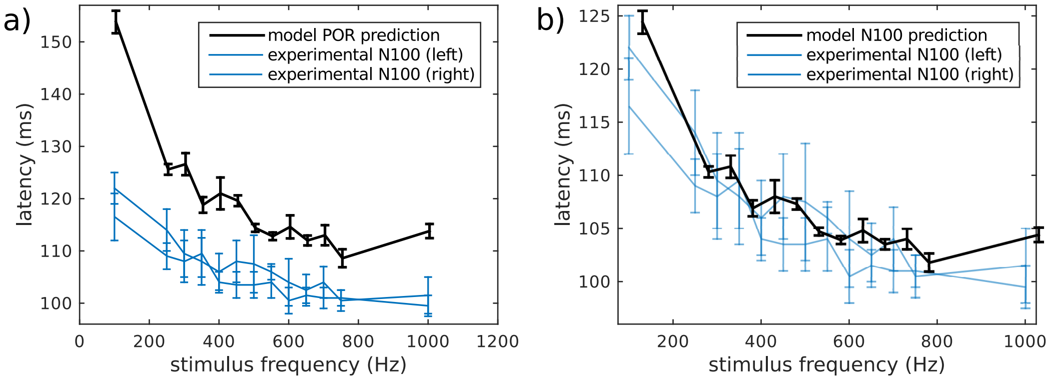


Figure 4.15: Predicted latencies for pure tones. a) Simulated POR latency values (black error bars) in comparison with N100 latency observations (blue error bars); the two experimental curves correspond to latency values observed in the right and left hemispheres. The large gap between both quantities is caused by the systematic bias introduced by the energy onset response in the N100 latency. b) Comparison of the corrected model predictions with the same experimental data. Predictions were averaged along 5 runs of the model; error bars are standard deviations. Experimental data was taken from [91], Fig 2.

.

Discussion

Cortical representations of pitch

Our cortical theory of pitch processing devises two separate networks holding subtly different representations of pitch: the decoder, located in the cortical generator of the POR; and the sustainer, located in the cortical generator of the pitch related SF. Both networks are

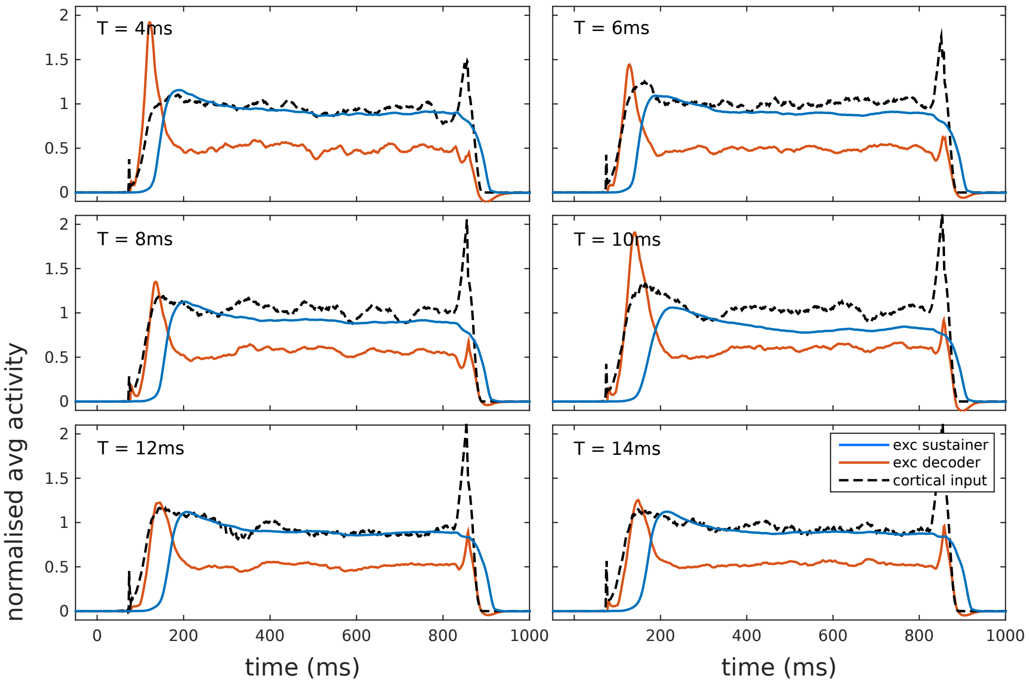


Figure 4.16: Sustained field morphology prediction for IRNs. Panels compare the average excitatory activity at the cortical input, the decoder layer, and at the sustainer layer. The three responses were normalised to an average activity of ~ 1 to improve the visualisation of the dynamics. The decoder's and sustainer's activity can be related to the POR's and SF's dipole moments (see §4.2.7.3), respectively. Note that, whilst the POR's dynamic respond rapidly to variations in the cortical input, an onset and offset delay characterise the SF's responses. Stimuli were IRNs with the same specifications as in §4.3.2. Simulations were carried out without cortical noise in this figure, in order to obtain a more accurate representation of the field dynamics.

putatively placed towards the lateral and anterior portion of the human Heschl's gyrus.

Inhibitory ensembles at the decoder layer activate only when there is enough information for pitch extraction available in cortex, and hence provide a robust pitch representation: a column n only becomes significantly active if the stimulus evokes the pitch corresponding to the of the characteristic period of the column. In contrast, excitatory ensembles at the decoder hold the subcortical representation during the first ~50 ms, and only after the decoding they present a robust pitch-rate representation of the evoked pitch. Pitch information in the decoder might be thus transmitted to higher cortical areas through feedforward inhibition [165], or through a combination of excitatory and inhibitory signals.

Responses at the sustainer layer are fully correlated with the perceived pitch; in exchange, they show a small delay with respect to the responses at the decoder. Whilst excitatory ensembles are pitch selective, inhibitory ensembles show the opposite behaviour: a population n only becomes inactive when the network encodes a pitch value $T = \delta t_n$. This reversed inhibitory representation might have a role in top-down modulation of subcortical areas [150] or in subcortical sensory integration [166], processes which are typically driven by selective inhibition. The representation at the sustainer's excitatory ensembles might be transmitted to higher cortical areas through feedforward excitatory signals. However, the sustainer's representation might be too slow to explain how can we experience an absence of pitch during short silence gaps that do not alter the sustained field (see §4.3.3).

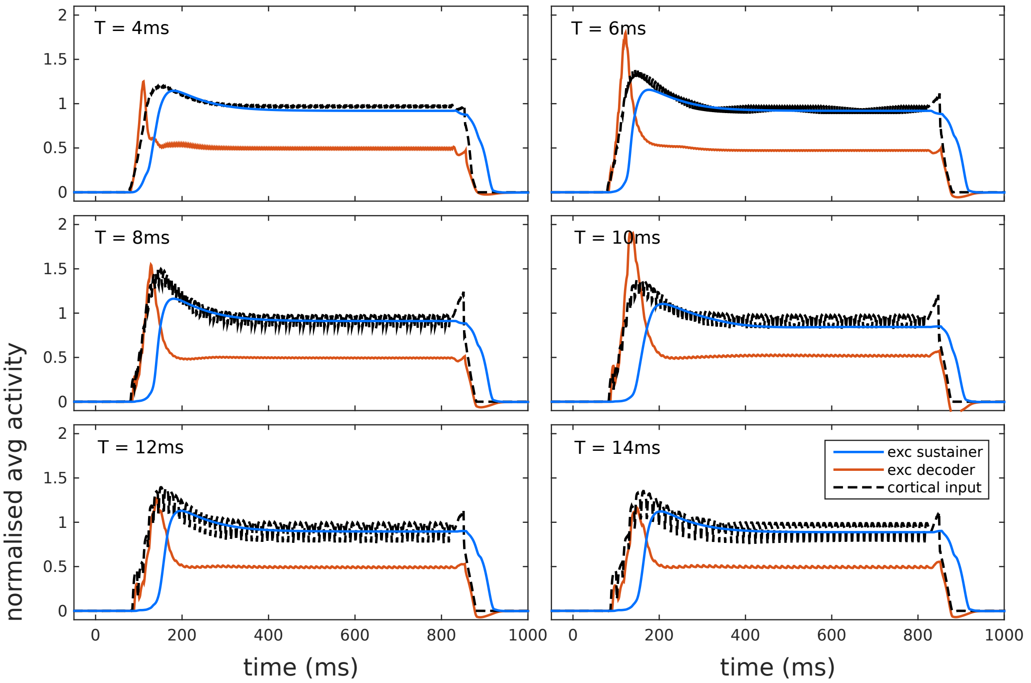


Figure 4.17: Sustained field morphology prediction for click trains. Plots compare the average excitatory activity in the cortical input, the decoder, and the sustainer. Stimuli were click trains with different inter-click intervals. Simulations were carried out without cortical noise in order to obtain a more accurate representation of the field dynamics.

A third neural representation, consisting of the harmonic patterns of activation associated to the regularised SACF, might be present in earlier cortical regions of the pitch processing hierarchy. So far, we have systematically identified this representation with subcortical processing, but there are several arguments in favour of a stable harmonic code permanently present in primary auditory cortex. First, harmonic co-activation of pitch-selective neurons is observed in cortex even after the tone's onset response [62,153], suggesting that the decoder is not the only cortical region holding the subcortical harmonic representation. Moreover, the coexistence of a pitch-rate representation and a stable representation of the spectral features of the SACF in cortex might explain the paradox of the spectral and fundamental listeners. This idea is explored later in §6.3.2.

The decoder and winner-take-all dynamics

The decoder model dynamics are reminiscent of the *winner-take-all* strategies typically used to represent neural processes in perceptual decision making [150, 151, 155]. Winner-take-all models typically consist of two (or more) competing excitatory populations, characterising the different possible outcomes of the perceptual decision, and a common inhibitory ensemble (see Figure 4.18). The strength of the thalamic input applied to each of the excitatory ensembles represents the amount of evidence available for each of the different perceptual outcomes. During the decision making procedure the excitatory ensembles are activated according to their relative likelihood; this activity then propagates to the inhibitory ensemble, which inhibits both excitatory populations equally acting as a mediator [155]. The decision making process eventually converges to a state in which one of the ensembles is totally shunted and the other one, representing the final decision of the system, remains active.

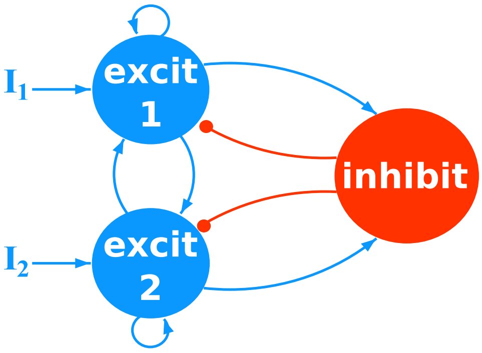


Figure 4.18: Schematic architecture of a winner-take-all system. Excitatory ensembles encode confronted perceptual outputs; the inhibitory ensemble mediates the decision. Figure adapted from [151], Fig 1.

Our system presents a behaviour of this kind, but on a much larger scale and with several added constraints. First, competition between ensembles is selective rather than global: only ensembles representing harmonically related frequencies compete with each other. Selective competition is encoded in the excitatory-to-inhibitory (directed towards higher harmonics) and inhibitory-to-excitatory (directed towards lower harmonics) connectivity weights (see §4.2.6). The common inhibitory ensemble arbitrating each competition is the inhibitory population corresponding to the higher common harmonic (i.e, the evoked pitch). This connectivity pattern might be the result of repeated co-activation of harmonically-related ensembles due to the structure of the SACF-like thalamic input.

Second, competition is biased towards the excitatory population located in the column of the inhibitory ensemble mediating the competition. This *bias* stems from the inhibitory-to-excitatory connectivity weights: excitatory ensembles in their same column are only weakly inhibited (see §4.2.6).

Neural competition at the decoder network mediates a template matching transformation and allows for concurrent pitch representations (see §5.3): simultaneous competitions from different harmonic shapes do not shunt each other, allowing for multiple pitch values being jointly represented in cortex. A side effect of the biased competition process is that a combination of two octaves (i.e, two tones with harmonically related fundamental frequencies) elicits a single pitch value, corresponding to the pitch of the higher note. Thus, our model provides a mechanistic explanation of the pitch of harmonic complex tones based on selectivity cortical dynamics.

Cortical dynamics of pitch processing

Decoding layer dynamics

Equations of the decoding network listed in §4.2.5 define a dynamical system with three variables per each excitatory ensemble (H^e, S^{AMPA}, and S^{NMDA}) and two per each inhibitory ensemble (H^i and S^{GABA}); i.e, a total of $5N = 1000$ dynamic variables termed here for simplicity $\vec{x}$.

Dynamic variables span a dynamical system whose behaviour characterises the model's dynamics. In absence of thalamic drive, the system presents a single state of stable equilibrium around the origin ($\vec{x} = \vec{x}_0 \simeq 0$).

Non-zero thalamic inputs change the stability properties of the system. An excitatory thalamic input moves the state of equilibrium towards a new attractor state termed here

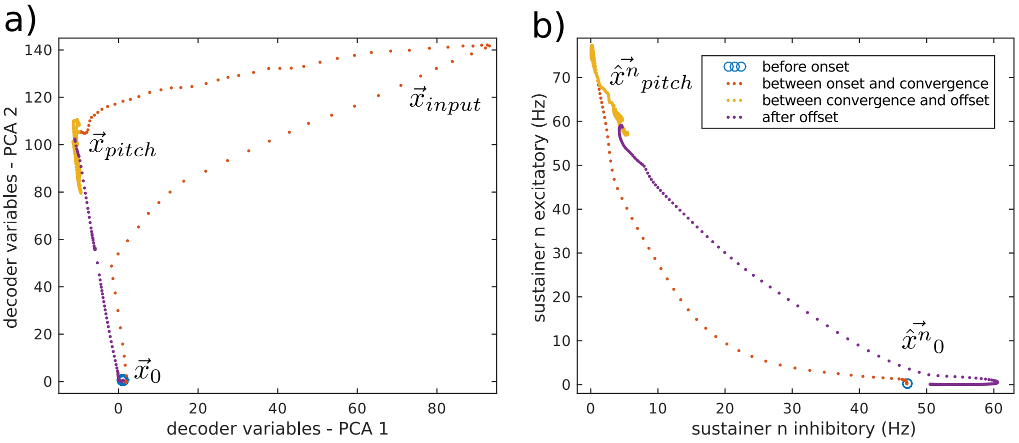


Figure 4.19: Attractor dynamics underlying pitch processing. Each dot represents the state of the system in an instant t, colours were used to characterise the different stages of the dynamics: open blue circles represent the absence of input (points are too close to each other to be distinguished); red dots represent states within the time window spanning from the stimulus onset to the convergence of the model to a specific pitch value (at about 100 ms after tone's onset); yellow dots represent states within temporal windows spanning from the convergence of the system to the tone's offset; purple dots show states in the time window corresponding to the the *relaxation dynamics* (see main text), spanning form the offset of the tone up to 200 ms after tone's offset. a) View of the most relevant dimensions of the decoder $\vec{x}$ during the processing of an IRN; dimensions were reduced using principal component analysis. The trajectory in the reduced space reveals key aspects of the onset and relaxation dynamics. The transition from $\vec{x}_0$ to $\vec{x}_{\text{pitch}}$ characterises the POR. b) View of the two dimensions of the subsystem characterising the decoded pitch n in the sustainer network. Note that the relaxation dynamics of the sustainer, corresponding to the transition from $\hat{\vec{x}}^n_{\text{pitch}}$ to $\hat{\vec{x}}^n_0$, are much slower than the relaxation dynamics of the decoder; this is characteristic of the sustained field offset delay (see text).

$\vec{x}_{\text{input}}$, where the excitatory populations represent the input activity (see Figure 4.19A). If the input presents a harmonic structure, a second equilibrium state we termed $\vec{x}_{\text{pitch}}$ (see Figure 4.19A), characterised by excitatory and inhibitory activation at the column encoding the pitch, arises in the system. Model dynamics transit from the initial state $\vec{x}_0$, to the input state $\vec{x}_{\text{input}}$, and then oscillate between this last state and $\vec{x}_{\text{pitch}}$.

We identify the POR as the neuromagnetic fingerprint of this two-stage transition: the build up of the transient corresponds to the transition $\vec{x}_0 \to \vec{x}_{\text{input}}$; the POR peaks shortly after the onset of the decoding process, characterised by the transition $\vec{x}_{\text{input}} \to \vec{x}_{\text{pitch}}$ (see Figure 4.19A). This identification allows us to connect the POR latency with the time necessary to trigger the $\vec{x}_{\text{input}} \to \vec{x}_{\text{pitch}}$ transition; thus, the POR latency reflects pitch processing time.

Sustainer layer dynamics

The role of the sustainer network is to modulate the dynamic properties of the decoder in order to prevent the reversed transition $\vec{x}_{\text{pitch}} \to \vec{x}_{\text{input}}$ and subsequent oscillations (see §4.2.4.1).

The sustainer's dynamics are much simpler than the decoder's and can be subdivided in $N = 250$ independent dynamical systems, one per column, with 5 variables each $\hat{x}^n$. At rest, the decoder's columns lie in equilibrium states $\hat{\vec{x}}^n = \hat{\vec{x}}_0$ characterised by a strong activation at the inhibitory population and null activation at the excitatory ensemble.

Under combined excitatory and inhibitory input from the decoder at a given subsystem $\vec{x}^n$, inhibition drops and excitation rises, switching to a new state $\vec{x}_{\text{sus}}$ termed here *sustained state* (see Figure 4.19B). Top-down efferents from the sustainer then lock the selective dynamics of the layer decoder, strengthening the attractor properties of the state $\vec{x}_{\text{pitch}}$ and turning it to a state of stable equilibrium (see Figure 4.19B).

Thus, we identify the pitch related sustained field as the fingerprint of the sustaining activation: the PSF activates when the sustained state of a sustainer subsystem is switched on, and deactivates whenever it is switched off. The dynamics explains the late onset of the field and the corresponding offset delay.

Relaxation dynamics

When the thalamic input is switched off, excitatory activity in the decoder drops, removing the excitatory input at the sustainer column $\vec{x}^n$, which returns to its resting state $\vec{x}_0$. As a result, the sustainer column stops modulating the dynamics at the decoder and the state $\vec{x}_{\text{pitch}}$ becomes, once again, unstable. Thus, the decoder state slowly relaxes back to the origin state $\vec{x}_0$.

If, after a period of silence, the same tone is played once again, the decoding process is re-triggered only if the duration of the gap is large enough as to allow the decoder to leave the state $\vec{x}_{\text{pitch}}$. Otherwise, the residual inhibition at the decoder prevents the peaks of the higher harmonics to rise again and keeps the inhibition at the sustainer column low, allowing it to transit back to the sustaining state.

Pitch transitions

A similar process drives the transition dynamics between pitch representations in sequences of consecutive notes eliciting different pitch sensations. A change in the harmonic structure of the cortical input induces a reaction in the cortical system, that triggers a new decoding process (see Figure 4.20).

Pitch changes can be understood as a reconfiguration in the attractor dynamics that drives the decoder from $\vec{x}_{\text{pitch}}$ to a new state $\vec{x}_{\text{pitch'}}$ (see Figure 4.21A). Since sustainer columns are independent of each other, dynamics are not altered by previous or posterior pitch representations (see Figures 4.19B and C). See also Appendix B for a more detailed discussion.

Conclusion

In this Chapter we have introduced a novel theory of cortical pitch processing based on selective neural competition. Our model fills the gap between the harmonic representations postulated by the temporal models of pitch perception (see §3.2.2 and §3.2.3) and the receptive-field-like representation reported in intracranial recordings in mammals (see §2.1.3.3); moreover, our theory postulates that the evoked field associated with the POR is a result of the transformation between these two representations.

The model describes cortical pitch decoding as a discrete event triggered by a sudden change in the subcortical input. Once a pitch value has been extracted from the subcortical representation, the decoded value is simply sustained (rather than being repeatedly decoded) until a new change in the subcortical input triggers a new decoding process. The model postulates that the pitch-related sustained field is associated to the sustaining process.

The sustaining strategy is reminiscent of *predictive coding* [162, 168] and reversed hierarchical strategies [21, 31], where top-down efferents transmit expectations about the input (in our model, expectations about the harmonic structure of the input), whereas bottom-up

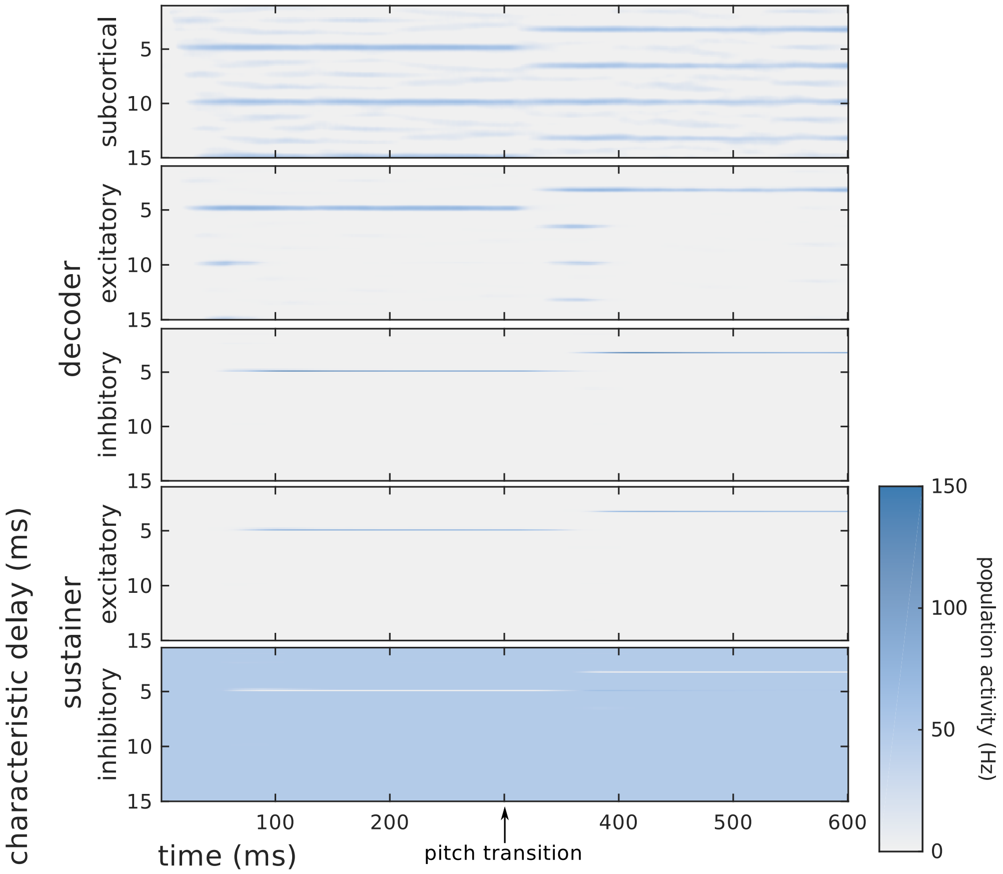


Figure 4.20: System's reaction to pitch changes. Response of the cortical system to pitch changes. Stimuli were IRNs with the same specifications as in §4.3.2.1; first tone had a fundamental frequency $f_0 = 200\,\text{Hz}$, second tone was its perfect fifth, with $f_0 = 300\,\text{Hz}$. Transition occurs 300 ms after the onset of the first tone (see arrow in the figure).

afferents carry prediction error (in our model, connections from the decoder network reinforce the corresponding sustainer column as long as the sustainer's expectations coincide with the decoder's input) [21, 169].

Expectations based on prior knowledge or experience could be used to modulate the decoding dynamics by reducing the inhibition at the sustainer column characterising the expected pitch, which facilitates the sustainer read-out and favours the selected pitch through the top-down excitation-to-inhibition efferents.

Decoding dynamics of the model quantitatively explain the dependence of the POR latency with four times the period of the stimuli: the cortical model needs three peaks of the harmonic series in order to robustly identify the harmonic pattern with the elicited pitch value. Although previous studies in pitch perception had postulated before that pitch processing requires several period cycles in order to make a perceptual decision [22, 43, 107], the mechanism of such integration was still unclear. Moreover, although MEG recordings (e.g [22]) show that the first physiological response integrating along several repetitions cycles is located in cortex, phase-locked activity is not robustly present above 50–200 Hz in Heschl's Gyrus [16]; thus, evidence for a periodic behaviour at several different cycles of the fundamental period of the stimuli should be transmitted to cortex from subcortical regions in order to explain the perception of pitch in medium and high frequency tones.

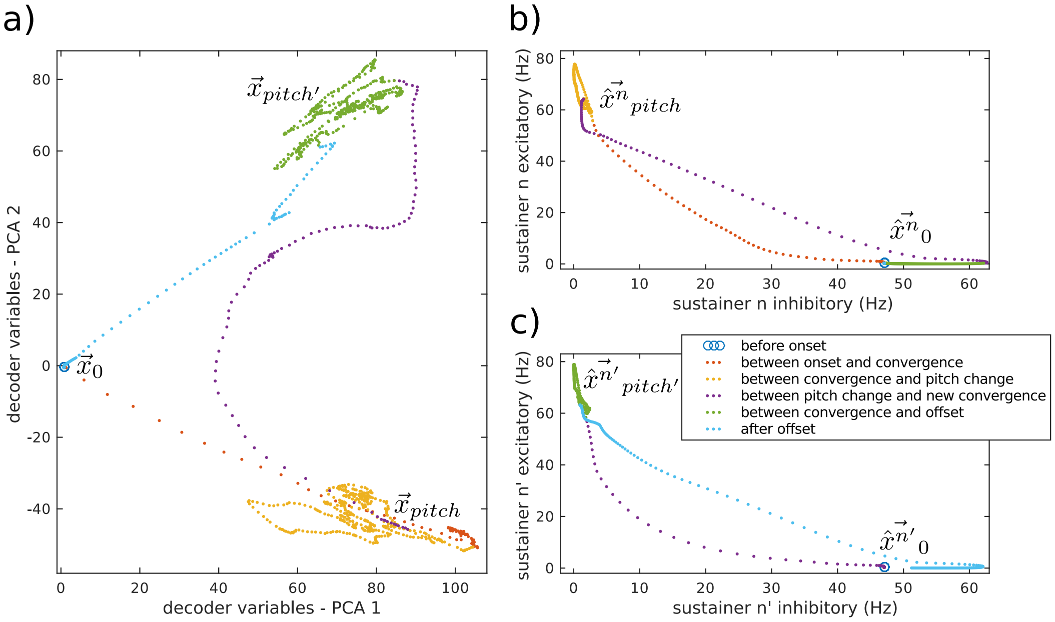


Figure 4.21: Attractor dynamics under pitch transitions. As in Figure 4.21, each dot represent the state of the system in an instant t. Two colours were added to represent the new stages in the system's evolution. Purple now represents the dynamics from the second stimulus onset to the new state of convergence, defined here as the state achieved 100 ms after the onset; green represents states between 100 ms and the second stimulus' offset; and light blue represent the states during the relaxation dynamics after offset. a) View of the most relevant dimensions of the decoder $\vec{x}$ during the processing of a sequence of two IRNs; dimensions were reduced using principal component analysis. Note that the transition from $\vec{x}_{\text{pitch}}$ to $\vec{x}_{\text{pitch'}}$ elicits a new POR, corresponding to the second stimulus. b) and c) View of the two dimensions of the sustainer columns characterising the first (b) and second (c) decoded pitch values. The first column relaxes back to the state $\hat{\vec{x}}^n_0$ after the pitch transition; similarly, the second column does not respond to the first tone, and only abandons the state $\hat{\vec{x}}^{n'}_0$ after the transition. Stimulation was the same as in Figure 4.20.

The harmonic representations provided by SACF-like models proposes an elegant solution for this problem.

Sustaining dynamics may explain several puzzling aspects of the sustained field behaviour. The late onset of the component is a consequence of the functional role of the sustainer, which only affects the processing dynamics after pitch has been extracted from the subcortical representation. The delay offset of the SF is a consequence of the sustainer's inertia, driven by the relaxation time constants of the sustainer's columns circuits; this inertia is crucial to prevent the system from constantly re-decoding the stimulus under noisy inputs [107].

Chapter 5

Consonance perception in vertical pitch interactions

The perception of music rests on synergistic interactions that produce new auditory objects whose identity is not simply provided by the sum of its parts. Interactions in the pitch dimension are of special interest in music, since they are responsible for the emergence of melody and harmony.

Pitch interactions are often studied on the basis of two orthogonal dimensions: consecutive tones in a melodic sequence are said to interact horizontally, whilst simultaneous tones in a chord or a dyad interact vertically [170]. As revealed by a landmark study in a subject with bilateral lesions in auditory cortex [171], horizontal interactions are mediated by plastic cortical areas responsible for high-level cognitive processing, and seem to be associated with a systematic exposure to certain musical intervals [172, 173].

In contrast, the substrate of vertical interaction might rely, at least partially, on intrinsic properties of pitch processing mechanisms located at different stages of the auditory pathway [171, 174–177].

This chapter explores the neurophysiological substrate of vertical pitch interaction; specifically, we will investigate the emergence of the sensations of consonance and dissonance from a pitch processing perspective.

The sensations of consonance and dissonance

Chromatic intervals

Chroma and pitch classes Harmonically related tones evoke different pitch sensations that are judged as similar by the listeners. When the tones are presented together, the resulting harmonic complex tone elicits the pitch of the tone with the lowest frequency (see §2.2.3.1). Harmonically related tones are said to share the same pitch class or *chroma* [5].

The chromatic scale The chromatic scale identifies 12 different pitch classes within the frequency range spanned from a given f_0 to the next superior harmonic $2f_0$ [4]. Pitch classes are defined by considering increasingly complex frequency ratios, which often elicit increasingly unpleasant sensations when interacting vertically forming *dyads* [4].

The simplest of such ratios defines the 7th pitch class or *perfect fifth*, characterised by $f_7 = 3/2\, f_0$. Dyads comprising a fundamental f_0 and its perfect fifth f_7 are generally judged as consonant, pleasant, sounds [4]. The most complex ratio considered in the Western scale

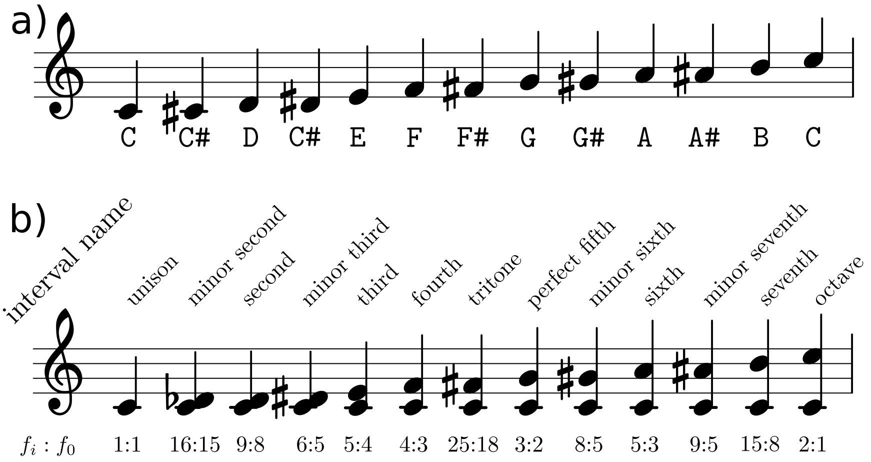


Figure 5.1: Chromatic scale and standard Western intervals. a) Chromatic scale beginning in `C`. b) First 13 standard Western intervals with baseline `C` and frequency ratios according to the natural tuning.

is the first pitch class or *minor second*; dyads comprising a fundamental f_0 and its minor second f_1 are often judged as dissonant, unpleasant sounds [4]. Figure 5.1B lists the first 13 intervals of the chromatic scale (f_0, f_n) sorted by ascending pitch value.

In Western music, it is common to use letters to denominate the different pitch classes and generate the chromatic scale starting with a baseline frequency f_0 near 440 Hz, denoted as `A` [4]. Subsequent frequencies in the chromatic scale are denoted by the symbols: `A#`, `B`, `C`, `C#`, `D`, `D#`, `E`, `F`, `F#`, `G`, and `G#`, after which the scale reaches the first harmonic of the series and its once again labelled as `A`. Numbers can be used to differentiate tones with the same pitch class along different octaves: `A3` often characterises $f_0 = 440\,\text{Hz}$, `A4` represents $f_0 = 880\,\text{Hz}$, etc. In music notation, the chromatic scale is written using ascending figures in a staff; for instance, Figure 5.1A shows a chromatic scale beginning at `C`.

Tuning systems There are several different tuning systems spanning a chromatic scale starting at a given f_0. Figure 5.1B shows the standard frequency ratios described by the *pure intonation*. Whilst this tuning schema ensues strong consonant sensations within the same octave, it fails to preserve harmonicity across different octaves for pitch classes different than the baseline of the tuning (e.g. in an intonation tuned to A4, the frequency $f_{\texttt{C4}}$ characterising `C4` deviates from $f_{\texttt{C5}} \equiv 2f_{\texttt{C4}}$ by a factor 80:81). The *equal tempered intonation* subdivides the chromatic scale in 12 segments according to the formula $f_n = f_0 * 2^{n/12}$ [4], guaranteeing that notes belonging to a given pitch class are perfect harmonics of each other across several octaves.

The sensation of consonance and dissonance

Dyads consisting of tones with different pitch classes elicit a new emergent sensation known as *tint* [178], commonly described as some degree of *consonance* or *dissonance*. Although consonance is often defined as a pleasant sensation and dissonance as an unpleasant one, we will argue that a clear distinction should be drawn between consonance and pleasantness.

The degree of consonance elicited by a tonal dyad depends mostly on the chroma relationship between the two tones, although it can be modulated by their pitch value [179]; for instance, although a perfect fifth is always more consonant than a minor second, perfect fifths are considered as more consonant in pitch regions over 440 Hz, whilst minor seconds

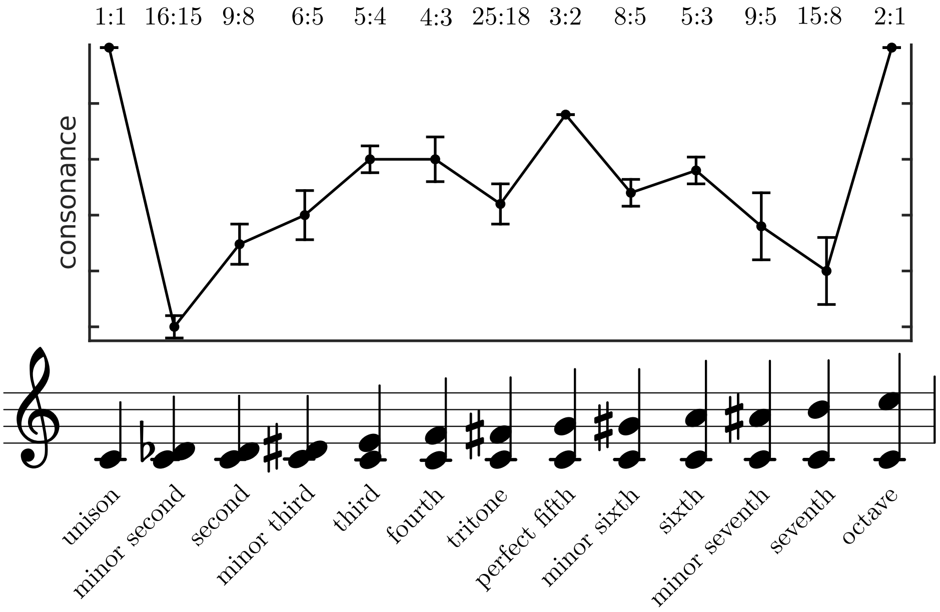


Figure 5.2: Typical consonance and dissonance judgements in dyads of harmonic complex tones. Consonance was calculated as the normalised averaged judgement across eight different studies (see Figure 6 in [180]).

are considered more dissonant in the pitch region spanning from 220 Hz to 7 kHz. Sounds' timbre and loudness do not significantly influence relative consonance judgements between intervals [179].

Consonance seems to be related to the simplicity of the integer frequency ratios of the tones comprising the dyads [172]; see Figure 5.2. This relationship leads to the classical relationship of consonance with *fusion*, hypothesising that dyads are more consonant the more they resemble a single tone [178].

Neural representations of consonance and dissonance

Frequency following responses elicited by dyads in human thalamus (see §2.3.2.1) show that neural activity in brainstem is phase-locked to the two tones present in the dyads, and that the robustness of the synchronisation was stronger in consonant combinations [174, 176, 181, 182].

EEG recordings report influences of consonance and dissonance over onset transient responses in auditory cortex [177], including the P30 complex [183], the N100 [175, 183, 184], and the P200 [184]. Moreover, Bidelman an colleagues. [175] found that the POR depth (see §2.3.2.4) is significantly modulated by consonance, indicating that consonance and dissonance are, at least partially, processed by cortical pitch-related mechanisms. Accordingly, a subject with amusia due to localised bilateral lesions in auditory cortex, able to capture the modal mood of musical excerpts, was unable to distinguish consonant from dissonant dyads [171].

Consonance is further represented in cortical regions typically associated with higher level cognitive functions such as superior temporal gyrus [171] and sections of frontal cortex [183]. The difference between the response to consonant and dissonant dyads in frontal areas is stronger in the right hemisphere [183].

The effect of musical training and cultural background

The role of neural plasticity on the perception of consonance and dissonance is still under debate: some authors argue that the sensation emerges naturally from the neurophysiology the auditory pathway, whilst other authors maintain that consonance perception is an acquired skill developed by cultural exposure [185]. We will argue that this controversy may be partly sourced in the erroneous equivalence between pleasantness and consonance sometimes drawn in the literature [172, 186], suggesting that tint is elicited by a general, culture-independent mechanism [172, 174, 187].

The influence of culture on perceptual judgements Cross-cultural experiments seem to indicate a general agreement in the judgement of consonance and dissonance in dyads. For instance, a study using subjects from USA, Germany and Japan, involving several generations, ages, and musical background, failed to find significant differences between the judgements drawn by different groups [185]. Similar conclusions were reported in another study comparing perceptual judgements from experts and naive subjects from Japanese and American backgrounds [187].

These results have been generalised to English, Farsi, Mandarin and Tamil speakers in a theoretical analysis considering the occurrence distribution of frequency ratios within the oral expressions of each language [180]. Occurrence maxima and minima were consistent along the four languages, and predictive of the chromatic scale's consonance and dissonance ratios in the natural tuning [180].

Consonance and pleasantness Some authors argue that a difference should be drawn between the concept of consonance and its associated pleasantness [172, 186]. An experiment comparing these two sensations in musicians and non-musicians found that, although non-musician subjects seem to judge consonant dyads as pleasant and dissonant dyads as unpleasant, there was a much smaller correlation between the pleasantness and consonance reported by experienced musicians [172]. Whilst pleasantness was judged differently by the two groups, the reported consonance was common to musicians and non-musicians [172].

Moreover, the pleasantness sensation elicited by consonant pitch combinations seems to depend strongly on cultural background: a recent study using subjects from the Tsimane', a society with minimal exposure to Western culture, showed no significant preference to consonant over dissonant dyads [186]. Although consonance preference had been previously established in 2- and 4-month-old infants [188], these results have been shown to reflect preferences built over a short-time familiarity rather than an intrinsic preference for consonant tone combinations [189].

Whilst pleasantness has an aesthetic connotation and presents large variances across subjects, depending on style and individual taste, there seems to be a general across-subject agreement on the relative consonance elicited by different intervals [178].

These studies seem to converge in the idea that consonant dyads are not universally perceived as pleasant or unpleasant, but that such association emerges by cultural exposure to determined musical styles. Reversely, the universality of the perception of consonance and dissonance across cultures suggest that tint is not a product of musical training, but a rather general biological phenomenon [172, 187].

Electrophysiological effects Although the phase-locking reliability of brainstem's FFR responses is stronger in subjects with musical experience, differences between the responses to consonant and dissonant dyads are not significantly modulated by training [174].

At the cortical level, responses depend on musical experience. However, the effects are much more robust in transients typically associated with higher-level cognitive functions, along the 300-800 ms latency band [177], than in the early onset transients elicited in auditory cortex, where differences are inconsistently reported across studies [173, 177, 184].

In summary, evidence converges in that, although cultural background and musical experience can enhance the sensation of consonance, its neurophysiological basis seems to be universal when controlling for pleasantness [176, 190].

Linear theories of consonance

In contrast to single tone's pitch, loudness, and timbre, the physical correlate of consonance in dyads is unclear. A general agreement exists that consonant dyads are characterised by simple frequency ratios [4], but this notion is far from producing robust systematic predictions of the level of consonance elicited by different sounds. For instance, dyads in lower frequency ranges generally produce smoother consonance/dissonance sensations than dyads in medium frequency ranges [172].

In this and the following section, we will review several families of consonance theories, to conclude that consonance is closely related to pitch processing.

Roughness theory

Early theories of consonance are based on the concept of *roughness*, described first by Helmholtz as a rapid beating sensation produced by dissonant tone combinations [4]. Helmholtz's theory suggests that dissonance results from the interaction between higher harmonics generally present in the tones comprising the chord [4]. Tone combinations derived from simple frequency ratios result in well-spaced higher harmonics that do not produce strong beating sensations.

A more recent reformulation introduced an abstract notion of roughness, as produced by interactions within a set of virtual tones called *difference tones*, arising from recursive additive operations between then tones comprising the chord [191]. Consistently, dissonant dyads elicited phase-locked activity synchronised with the frequency of the first order difference tone in monkey's primary auditory cortex [192]. Similar findings have been reported in auditory evoked potentials in human's Heschl's gyrus [192].

Although these findings could be consistent with a temporal theory of consonance, such theory is unable to explain why dyads with difference tone's frequencies over the cortical phase-locked limit elicit a consonance percept [192]. Moreover, roughness theory fails to explain why intervals beyond the octave are judged as consonant or dissonant as their within-octave counterparts [192].

The critical bandwidth theory

Dyads of pure tones are perceived as consonant when the tones' frequency separation exceeds a certain *critical bandwidth*, roughly equal to three semitones around the frequency of the fundamental tone [172], and dissonant otherwise [172]. The critical bandwidth theory suggests that dissonance is the result of such interactions between pure tones comprising the harmonic complexes typically observed in natural sounds. Theory's predictions show a good agreement with experimental data in dyads [193] and chords [179, 194] within the same octave.

The critical band theory is consistent with the roughness theory: dissonant intervals within the critical bandwidth produce rapid beating sensations, in agreement with Helmholtz's description of dissonance [172]. In fact, Helmholtz predicted that roughness was a consequence of the interaction between the partials of the harmonic complexes, although he assumed a fixed critical bandwidth of interaction [4].

Pitch processing theories of consonance

Evidence against the critical band theory

Early experiments testing the critical band theory intentionally omitted tritone and fourths because they were recognisable intervals and thus cultural background could affect consonance judgements [185]. This constraint severely limits the evaluation of the theory with pure tones, since most of the dyads over the three semitones are consonant, except for the tritone and the minor seventh.

Another important point of conflict is the dissonance percept elicited by dyads made of iterated rippled noises [175]. IRNs cannot be expressed as sums of pure tones and thus, according to the critical band theory, IRN dyads should not elicit a dissonant sensation at all. However, IRN dyads are judged as consonant or dissonant depending on the same fundamental frequency ratios as harmonic complex tones [175].

Consonance and pitch processing

A key study addressed the problem of consonance and roughness by systematically analysing the responses or 250 subjects to different dyads through three different dimensions: preference to consonant over dissonant dyads, preference to dyads eliciting a rough or beating sensation, and preference to dyads whose spectral shapes matched in greater or less degree a harmonic series [190]. The study revealed that, whilst listeners generally showed preference for both dyads not eliciting a beating or rough sensation and dyads whose spectra were closest to a harmonic series; only the spectral criterion was consistently correlated with a preference for consonant over dissonant dyads [190]. This important result suggests that the physical correlate of consonance and dissonance relies on harmonic relationships rather than on the elicited roughness [190].

Accordingly, subjects with impaired pitch perception have been shown to be unable to discriminate between consonant and dissonant dyads, even though they were sensitive to beating and roughness [185, 195], suggesting that pitch processing plays a crucial role in consonance and dissonance perception. The hypothesis is further supported by EEG cortical recordings using noise-to-IRN stimulation, which show that the POR elicited by IRN dyads and located in the anterolateral portion of Heschl's gyrus was significantly modulated by the perceived consonance [175].

Consonance and the SACF

A last link between consonance and pitch processing is provided by the fact that, whilst consonant dyads such as the fourth or the fifth are associated with periodic autocorrelation functions, dissonant dyads produce aperiodic SACF representations [185]. Consonance seems to be closely correlated with regularities in the temporal structure of the subcortical activity elicited by the sound at the same temporal scale of pitch processing [185, 196].

A series of studies in FFR by Bidelman and colleagues [176, 181, 182] compared the autocorrelation function of the brainstem frequency following responses elicited in human subjects by consonant and dissonant dyads. ACFs associated to consonant sounds had a larger degree of similarity with periodic templates than ACFs associated to dissonant sounds [176, 181, 182].

Discussion

The universality of consonance The universality of a particular preference for consonant or dissonant chords is still under debate: subjects from a large diversity of cultural backgrounds and different degrees of musical expertise seem to share a common preference for consonant combinations [180,185,187], but tribal groups with no exposure to Western culture [186] and 2–4-months-old infants show no particular preference for one or another [189]. However, the difference between the sensations of consonance and dissonance do seem to be of an universal nature [172, 174, 178, 180, 185, 187, 189], in the same way that loudness, timbre and pitch variations are universally perceived as intrinsic properties of single tones.

Consonance and the SACF Although the physical correlate of consonance is still a matter of controversy, the theory relating consonance to the regularity of the SACF of brainstem responses associated to a given pitch combination is the most compelling option among the current competing hypotheses [176, 181, 182, 185]. This theory links consonance perception to the classical notion of consonance as a result of *fusion* [196], that identified consonant chords as those most resembling a single tone [178].

In §3.5.2, we argued that the regularised SACF can approximate the subcortical representation of pitch, by processing the temporal structure of the auditory nerve activity in the temporal scales typically characterising pitch perception. Thus, we suggest here that the sensations of consonance and dissonance are characterised by these same temporal scales in pre-cortical stages.

Consonance and pitch processing Previous studies have shown that cortical pitch processing is an essential prerequisite in order to differentiate consonant from dissonant dyads [185,195]. Consonance could be processed after cortical pitch extraction: pitch values could be transformed into a chromatic representation and then compared through a neural network using some sort of consonance and dissonance template matching mechanism.

However, unlike pitch, that can be used to determine the size of the sound source or to disassociate sounds coming from different sources, consonance and dissonance do not seem to present a clear functionality justifying the existence of a dedicated decoding mechanism. Moreover, EEG recordings report a strong correlation between POR amplitude and the degree of perceived consonance, emphasising the role of the putative pitch centre on consonance processing [175]. Taken together, these observations suggest that consonance might be a collateral effect of pitch processing.

In the next sections, we will explore this last hypothesis combining novel MEG recordings on IRN dyads and the cortical model of pitch processing developed in Chapter 4. Our results provide further evidence suggesting a causal link between cortical pitch processing and the emergence of consonance and dissonance sensations.

AEF dynamics during consonance processing

Previous studies considering the POR dynamics

Using a noise-to-IRN paradigm in an EEG experimental set, Bidelman and Grall [175] measured the POR associated to the onset of dyads eliciting different degrees of consonance and dissonance. They found a significant positive correlation between reported consonance and the POR's depth, and a subtle non-significant negative correlation between consonance and the POR's latency [175].

Although Bidelman's results are highly revealing, their experimental conditions were designed to investigate the hierarchical organisation of consonance perception rather than the dynamics of consonance processing within auditory cortex. As a consequence, the experimental conditions could be improved in order to understand further the behaviour of the cortical dynamics associated to the onset of dyads.

First, an MEG rather than EEG paradigm could offer a greater spatial resolution and an increased signal-to-noise ratio of fields sourced in anterolateral Heschl's gyrus (see §2.3.1.3).

Moreover, in Bildelman's experimental setup, each note of the dyad was delivered to a different ear in order to avoid possible cochlear beating during the stimulation [175]. Binaural stimulation has been shown to produce stronger perceptual and physiological differences between the responses to consonant and dissonant dyads [197].

Experimental methodology

We designed an experimental setup in order to investigate the dynamics of POR for dyads in auditory cortex. Experimental conditions were here optimised to further investigate the dependence of the POR's latency with consonance, already hinted in Bidelman's study. Specifically, the study aimed to find a significant correlation between dissonance and the POR's latency, and measure the size of the effect in musicians and non-musicians. The study was carried out by Sebold, Andermann, and Rupp at the Department of Neurology, Heidelberg University, Germany.

Details on the stimuli Dyads consisted of two simultaneous 8-iteration IRNs sampled from the same white noise (Bidelman's IRNs were generated with 64 iterations, eliciting a stronger pitch sensation). The ground note in the dyads was set to $f_0 = 160\,Hz$ (Bildeman's ground IRN presented a higher pitch value, with $f_0 = 250\,\mathrm{Hz}$). Stimuli were presented diotically (Bidelman's study used dichotic dyads) and were preceded by a Gaussian noise in order to disentangle the POR from the energy onset response (see §2.3.2.4). Noise and dyad durations were set to 750 ms and the transition was mediated by a 10 ms Hamming window. We chose a lower number of iterations than Bidelman and diotic stimuli in order to maximise the effect of tint over pitch in the POR. A lower overall pitch was chosen in order to increase the latency of the onset response and boost the statistical power of potential latency differences.

Experimental paradigm and AEF recording Auditory evoked fields were recorded using a Neuromag 122 MEG system on 37 subjects. In comparison, Bidelman's study used a 64-channel EEG equipment and 9 subjects [175]. 250 trials were recorded per dyad; trials were pseudorandomised and separated by 1000 ms inter-stimulus-intervals.

Recording the responses to each dyad requires $250 \times (1500 + 1000)\,\mathrm{ms} = 10.4$ minutes. In order to guarantee that the subject was vigilant during the whole experiment, duration was constrained to approximately one hour, corresponding to six different stimulus conditions: three consonant (unison, third, and perfect fifth) and three dissonant dyads (minor second, tritone, and minor seventh); see Figure 5.4B. Bidelman's study recorded the responses to all 13 possible chromatic dyads.

Data analysis and preprocessing Data was preprocessed using standard MEG procedures (see, for instance, Methods in [1]). Equivalent dipoles were fitted for the POR and the sustained field for each subject and hemisphere over the pooled six conditions, based on the assumption that the field sources were the same in all dyads.

Experimental results

Grand averaged dipole moments at the pitch onset response and sustained field generators for each of the stimulus conditions are shown in Figure 5.3.

POR latency and depth dependence with consonance In agreement with Bidelman and Grall's study [175], consonant dyads elicited a larger ($p < .0001$; see Figure 5.4A) and earlier POR response ($p < .0001$; Figure 5.4B). In contrast with Bidelman's EEG

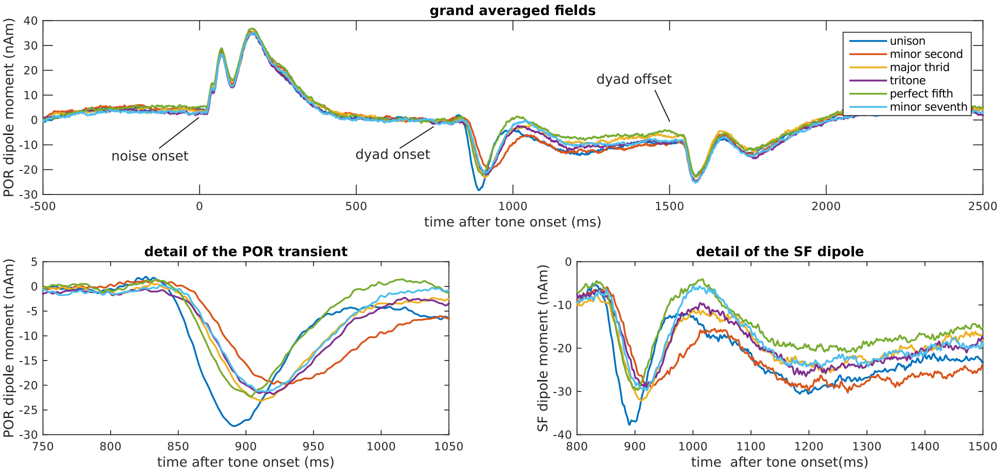


Figure 5.3: Averaged evoked fields elicited by consonant and dissonant dyads. a) Temporal waveform of the grand averaged dipole moment of the POR generator for each of the six dyads considered in the experiment. b) Detail of the POR triggered by the transition between the white noise and the IRN dyad. c) Grand averaged dipole moment at the sustained field generator.

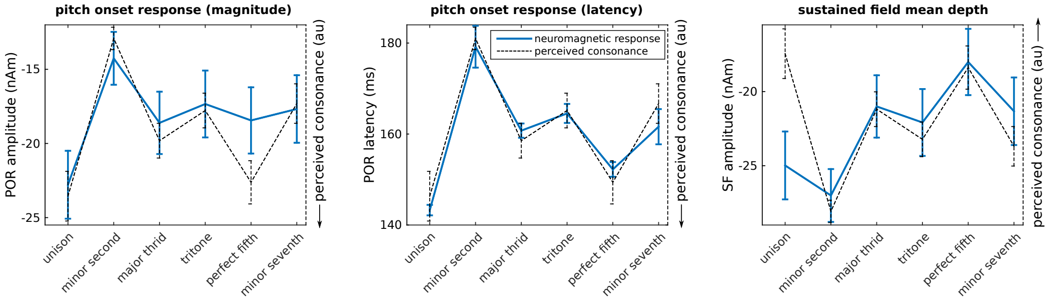


Figure 5.4: Comparison between the POR and SF properties and perceived consonance in IRN Dyads. a) Averaged POR amplitude. b) Averaged POR latency. c) Averaged depth of the sustained field. Error bars are standard errors; perceptual results were taken from [175], Fig. 2.

results, the latency of the magnetic POR showed a highly significant dependence with elicited consonance.

Sustained field magnitude dependence with consonance The magnitude of the elicited sustained field was smaller for the pooled consonant than for the pooled dissonant dyads ($p < .0001$). However, depth ordering did not generally mirror consonance judgements (see Figure 5.3c).

Lateralisation of the responses No significant differences were observed between the left and the right hemisphere on the POR morphology (latency: $p = .320$; amplitude: $p = .029$), on the SF amplitude ($p = .011$), nor between the responses to consonant or dissonant dyads (POR latency: $p = .489$; POR amplitude: $p = .334$, SF amplitude: $p = .417$). Significance analysis in this section were perfomed by M. Andermann using the bootstrap method [198].

Effects of musical training Individual musical aptitude of each subject was assessed using the Advanced Measures of Music Audition (AMMA); a standardized test that measures musical aptitude independently of expertise [199]). The median AMMA score of the 37 subjects was used to divide the sample into a low-AMMA group and a high-AMMA group.

No significant difference was found between the POR amplitude elicited by subjects in the high- and low-AMMA groups, however, the POR latency of was significantly shorter for the high-AMMA group ($p = .003$). Furthermore, the elicited sustained field was found to be significantly larger for the high-AMMA group ($p = .003$).

However, the difference between the responses to consonant or dissonant dyads did not vary between high- and low-AMMA listeners (latency: $p = .152$; amplitude: $p = .479$). No significant interactions were neither found between high- or low-AMMA scores and hemisphere activity (latency: $p = .012$; amplitude: $p = .358$).

Discussion

Experimental results show that the degree of consonance elicited by different pitch combinations has a direct and significant influence on the temporal dynamics of the pitch onset response.

Previous studies had already shown that pitch processing was an essential prerequisite for the emergence of the sensations of consonance and dissonance [185, 195]. Bidelman and Grall's study revealed that the magnitude of the POR was significantly modulated by consonance [175], although this effect might just reflect a decay in the robustness of phase-locking observed in the FFR in human brainstem for dissonant dyads [174, 176, 181, 182].

On the contrary, the effect of consonance on the POR's latency cannot be explained as a consequence of the FFR's fidelity decrease; for instance, the number of iterations in an IRN, although affecting the salience of the SACF representation of the tone and the elicited POR's amplitude, do not have an effect on the POR temporal dynamics [22]. The effect can neither be accounted for as a result of the POR's latency dependence with pitch typically reported in the literature; for instance, the latency shift expected between $f_0 = 160\,\text{Hz}$ and its minor second $f_0 \sim 170\,\text{Hz}$ is of $\sim 2\,\text{ms}$, whilst a $\sim 36\,\text{ms}$ latency gap is observed between the unison and the minor second dyads.

Thus, a more fundamental phenomenon seems to underlie the strong latency correlation found with perceived consonance. In the next section, we will investigate the neural substrate of such correlation using the cortical theory of pitch processing developed in Chapter 4.

Pitch processing during consonant and dissonant interactions

In this section, we will investigate the behaviour of pitch processing mechanisms presented in Chapter 4 when the input stimuli consists of consonant and dissonant dyads. Our results suggest that the POR latency dependence with consonance reported above (see §5.2.3) can be explained as a consequence of *harmonic collaboration* between the SACF subcortical patterns during the pitch decoding procedure. Moreover, we will argue that this process might be the underlying mechanism behind the emergence of the sensations of consonance and dissonance.

Subcortical processing of IRN dyads

The same procedure as in §4.2.2 was used to calculate the regularised SACF associated to the different dyads considered during the experimentation. Dyads consisted of two simultaneous

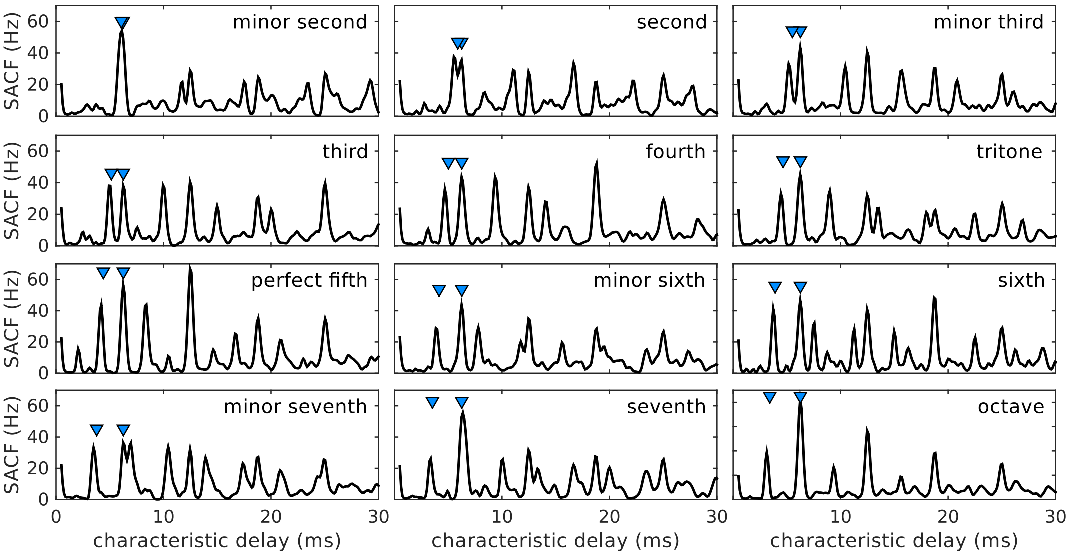


Figure 5.5: Averaged regularised SACF elicited by IRN dyads. Blue triangles point to the actual periods of the tones comprising the dyads. Note that SACF patterns associated to consonant dyads like the perfect fifth show a more periodic pattern that SACF associated to dissonant dyads like the minor second. Stimuli were IRN dyads as described in §5.2.2; plots show the regularised SACF according to the compensated rescaling factor $\mu_0 = 150\,\text{Hz}$.

IRN tones with 8 iterations, generated from the same Gaussian sample, and bandpassed between 125 Hz and 2 kHz. We used the same parameters as in the MEG experimentation described in §5.2.2 in order to reproduce the behaviour of the cortical mechanisms operating during the recordings in as much detail as possible.

We considered 13 dyads, corresponding to each of the notes of the chromatic scale (see §5.1.1), with the same ground tone $f_0 = 160\,\text{Hz}$ as in the MEG experimentation. Averaged regularised SACF for each dyad are shown in Figure 5.5. The resulting SACFs showed harmonic series with a remarkably lower signal-no-noise ratio than their single-pitch counterparts. The degradation of the SACF peaks is a consequence of the auditory nerve activity being simultaneously phase-locked to two different fundamental periods, effectively degrading the robustness of the phase-lock associated to each independent pitch.

In order to compensate for this effect, we doubled the rescaling factor $\mu_0 \to 2\mu_0 = 150\,Hz$, which yielded peaks of activation of $\sim 60\,\text{Hz}$, comparable to the activity elicited by simple IRNs. In §6.2.2.2 we will argue that an adaptive integrative mechanism might be responsible for an effecitive rescaling modulation of this kind.

Extracting simultaneous pitch values

Adjustment of the parametrisation

Next, we tested the cortical model described in §4.2 using the regularised SACF associated to the different dyads as subcortical inputs. At this stage, it was necessary to perform a fine tune of several parameters of the cortical model in order to allow for concurrent pitch representations in the decoder's readout. Prior to this tuning, the model failed to converge to any of the pitch percepts present in most of the dyads except for the minor second, where the model converged to an average of the pitch of the two IRNs.

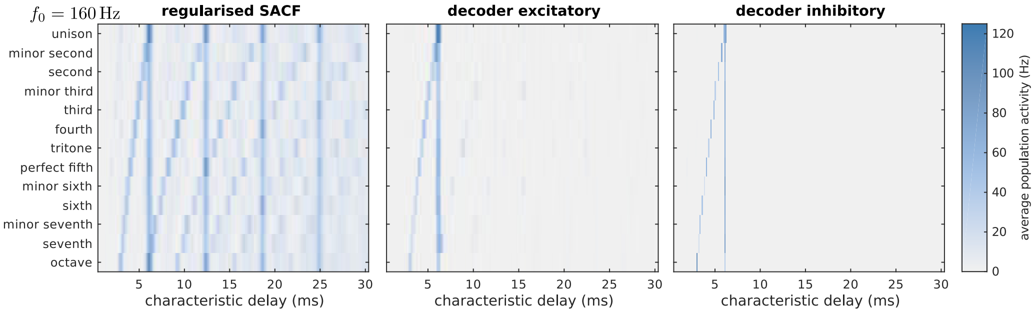


Figure 5.6: Perceptual predictions of the model for IRN dyads. Average responses of the activity of different model's populations elicited by IRN dyads; fundamental frequency of the lower note was $f_0 = 160\,\mathrm{Hz}$. Responses were averaged in time (as in Figures 4.7–4.11) and across five runs of the model to investigate trial-to-trial variability. As in § 4.3, each row shows the responses to each stimuli (in this case, to each dyad); note that, except for the unison, all dyads elicit the activation of two different pitch-selective populations, one corresponding to the pitch of the ground tone $f_0 = 160\,\mathrm{Hz}$ and one corresponding to the note of the chromatic scale characterising the dyad.

We tuned the thalamic conductivity J^{th}_{AMPA}, decoders' conductivities targeting inhibitory ensembles J^{ei}_{NDMA} and J^{ii}_{GABA}, afferent conductivities $\hat{J}^{a}_{\mathrm{AMPA}}$ and $\hat{J}^{a}_{\mathrm{GABA}}$, and the efferent conductivity $\hat{J}^{e}_{\mathrm{NMDA}}$. Parameters were adjusted to ensure that the model was able to decode both pitch values from the subcortical inputs of 6 representative dyads (see §A). The extraction of the two tones in the minor second was specially challenging because of the overlap of the peaks corresponding to their fundamental frequency in the SACF representation (see Figure 5.5B; see also §6.1.1.2A).

Parameters variation after tuning did not exceed more than 10% of the original value of the constants, and the final parametrisation of the model was consistent with all previous results. Parameters reported in Table 4.1; all results and Figures reported in §4.3–4.4 were produced using this last parametrisation.

Perceptual results

Model's representation of the IRN dyads after the decoding is shown in Figure 5.6A. Simultaneous tones were successfully extracted from the SACF representations in all tested dyads with ground fundamental frequency $f_0 = 160\,\mathrm{Hz}$.

Using the entire SACF pattern for pitch extraction is crucial to explain the perception of dyads of tones with similar fundamental frequencies. Consider for instance the minor second: the first two peaks of the harmonic series corresponding to the fundamental f_0 and its minor second overlap in a single peak corresponding to the mean characteristic lag of each of the two notes (see second raw of the left panel in Figure 5.6). The series are only independently resolved when looking at the subsequent harmonics. Note that the overlap of the peaks is still evident in the excitatory populations at the decoder layer, whilst the peaks are clearly resolved independently in the representation held by the decoder's inhibitory populations.

Decreasing the fundamental frequency of the lower tone in the dyad yields similar perceptual results (see Figure 5.7A). However, the tones of dyads starting at frequencies above $\sim 250\,\mathrm{Hz}$ are not always resolved by the cortical model (see the perceptual predictions for the sevenths in Figure 5.7B). This is due to the overlap of the first peaks of the harmonic series corresponding to each of the IRNs comprised in the dyads in the SACF representation.

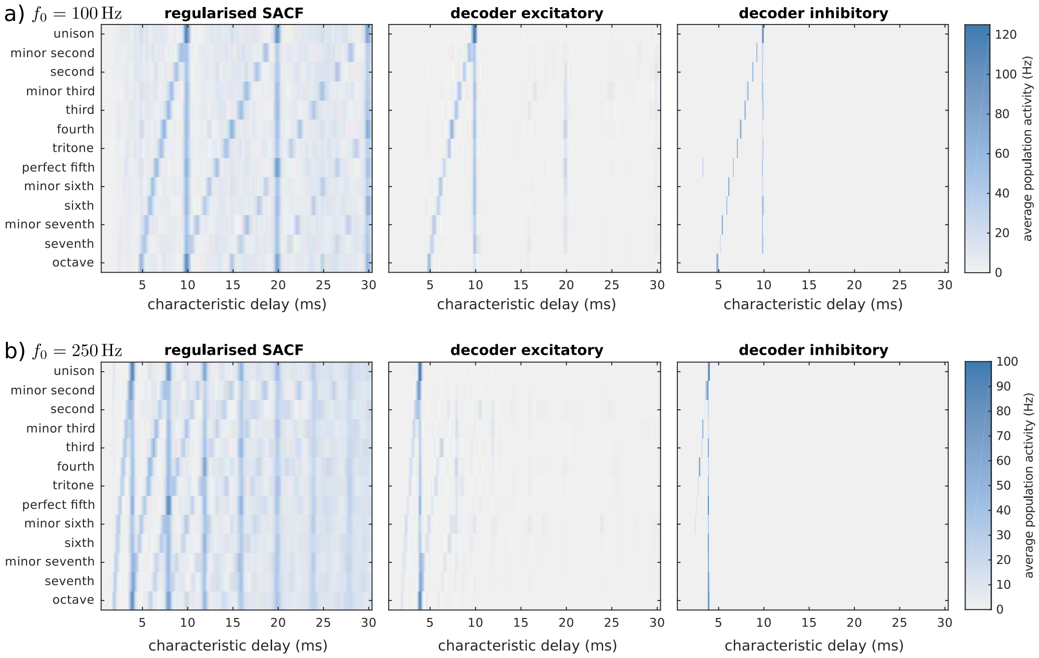


Figure 5.7: Perceptual predictions of the model for higher and lower-pitched IRN dyads. a) Average responses of the activity of different model's populations elicited by IRN dyads with lower fundamental frequency $f_0 = 100\,\mathrm{Hz}$. b) Responses for IRN dyads with lower fundamental frequency $f_0 = 250\,\mathrm{Hz}$. Note that, in this last case, some lines characterising the extracted pitch in the inhibitory ensembles are missing or under-represented. Methodology of the simulation was the same as in Figure 5.6.

A solution to this issue would be to allow for the model to integrate along a higher number of harmonics when the first three peaks of the series are not enough to perform a reliable perceptual decision (see §6.2.3.1).

The model was further tested using dyads based on HCTs and click trains. Perceptual results were not significantly different from the results obtained with IRN dyads (see Figure 5.8).

Decoding dynamics and evoked fields

Neuromagnetic fields elicited by the decoder network

Since the cortical model was able to extract a reliable representation of simultaneous pitch values in IRN dyads, next we inspected the derived evoked fields at the POR generator (see §4.2.7.2). A systematic 25 ms delay between the prediction and observed neuromagnetic trends was observed in all conditions; the delay was corrected by adjusting the cortical delay $\Delta \rightarrow \Delta_{\mathrm{dyads}} = \Delta + 25\,\mathrm{ms} = 95\,\mathrm{ms}$. This systematic increase of the cortical delay might be connected to the correction factor of the SACF: a decrease in the signal-to-noise ratio of the subcortical input typically yields an increase of the necessary processing time.

Figure 5.9 shows a comparison of the grand average of the equivalent dipole moment elicited by each dyad at the POR generator (same as in Figure 5.4A), in comparison with a single-trial field predictions of the model for the same stimuli after the adjustment of the cortical delay. Fairly good agreements with the morphology of the recorded fields are observed for all conditions, with the exception of the unison where the amplitude of the

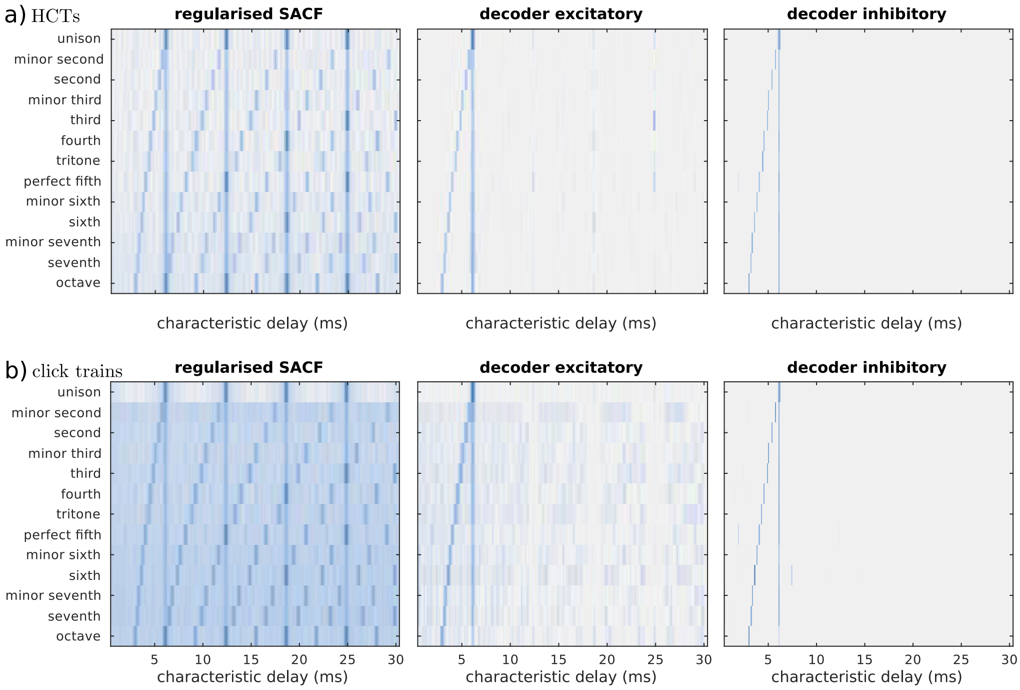


Figure 5.8: Perceptual predictions of the model for HCT and click train dyads. a) Average responses of the activity of different model's populations elicited by HCT dyads with a ground fundamental frequency of $f_0 = 160\,\text{Hz}$. b) Responses for click train dyads with an inter-click interval equivalent to $f_0 = 160\,\text{Hz}$. Methodology of the simulation was the same as in Figure 5.8.

POR transient was underestimated by the model.

POR's latency dependence with consonance

Next, we tested if the model was able to replicate the POR latency observations robustly across trials. Figure 5.10A compares the predicted POR latency with the MEG observations. A strong agreement between experimental and modelling results are observed for all conditions. Moreover, we found a strong correlation between perceived consonance and the predicted POR latency for the rest of the dyads, as shown in Figure 5.10B.

Our model explains the correlation between POR's latency and consonance through harmonic facilitation. Pairs of tones resulting in consonant combinations show obvious similarities in their SACF-like harmonic series: the SACFs associated to an $f_0 = 200\,\text{Hz}$ tone and its perfect fifth, an $f_0 = 300\,\text{Hz}$ tone, share one in three peaks in the harmonic series; in contrast, the SACFs associated to a $f_0 = 200\,\text{Hz}$ tone and its minor second do not share any peak in the entire range of the SACF under the frequency limits considered in our model. These coincidences effectively facilitate pitch processing and speed it up: common harmonics contribute to the build up of both tones, explaining why the perfect fifth or the fourth dyads are extracted considerably faster than dissonant dyads. This explanation is fully in line with the observation that consonant dyads produce more periodic, harmonic like, patterns (see also §5.1.6.3).

A second phenomenon, having a more subtle but also noticeable effect, is the slightly larger degradation of the signal-to-noise ratio observed in dissonant dyads. These differences are responsible for the more subtle differences observed between dissonant dyads.

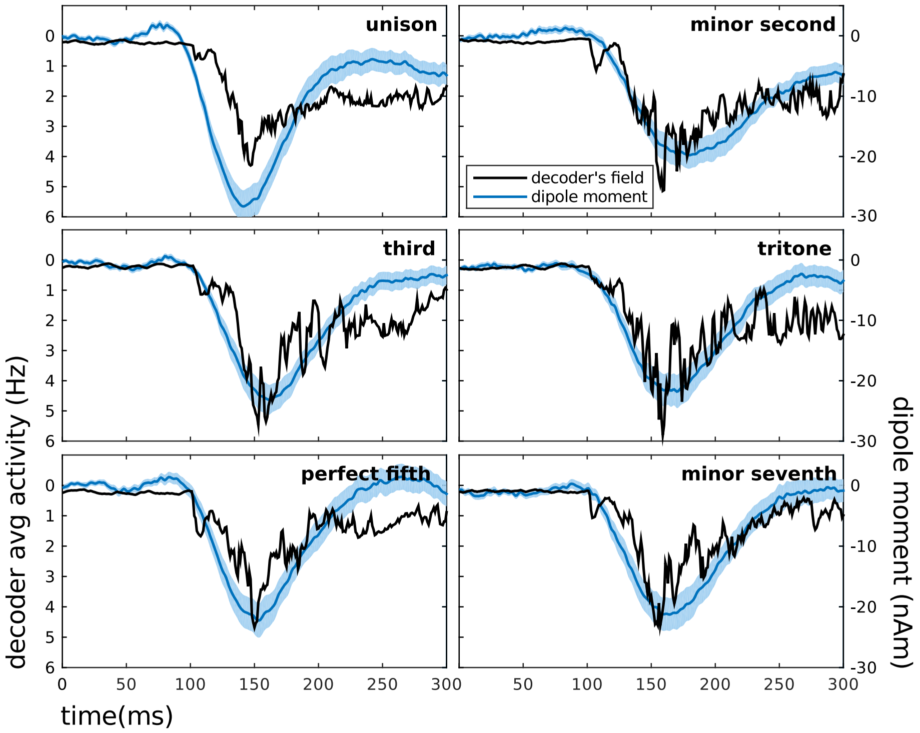


Figure 5.9: Comparison of the model's simulated fields and the POR dynamics for dyads. Plots compare the measured dipole moment (blue, shadows are standard errors; same as in Figure 5.3) with the average activity of the model's excitatory populations at the decoder (black), predictive of the POR neuromagnetic field. A very good agreement between both quantities is observed for all dyads except for the unison, where there is a significant difference between the measured and predicted amplitude of the POR. This divergence is due to a fundamental limitation of our model: the POR is computed as the aggregated activity across excitatory populations, which depends on the amount and shape of the peaks emerging in the SACF representation. Dyads with several pitch values elicit a longer number of harmonic peaks, eliciting a larger activity in the decoder's excitatory ensembles.

In order to corroborate the robustness of the correlation between the PORs latency and perceived consonance we computed the latency predictions for two extra families of dyads. Latency predictions for dyads of harmonic complex tones are displayed in Figure 5.11A; latency predictions for IRN dyads starting at a ground pitch $f_0 = 100\,\text{Hz}$ are shown in Figure 5.11B. Predictions for both families show consistent correlations with consonance.

Discussion

Modelling results suggest that the observed differences between the latency of the POR's elicited by consonant and dissonant dyads is caused by the collaborative interaction in cortex between the processed harmonic series characterising each of the two notes. This process is combined with the competitive interaction between populations described in the previous chapter; coalescing in the emergence of the consonance sensation. Thus, pitch processing of each one of the notes is not only a prerequisite, but it enables auditory cortex to elicit the perception of consonance.

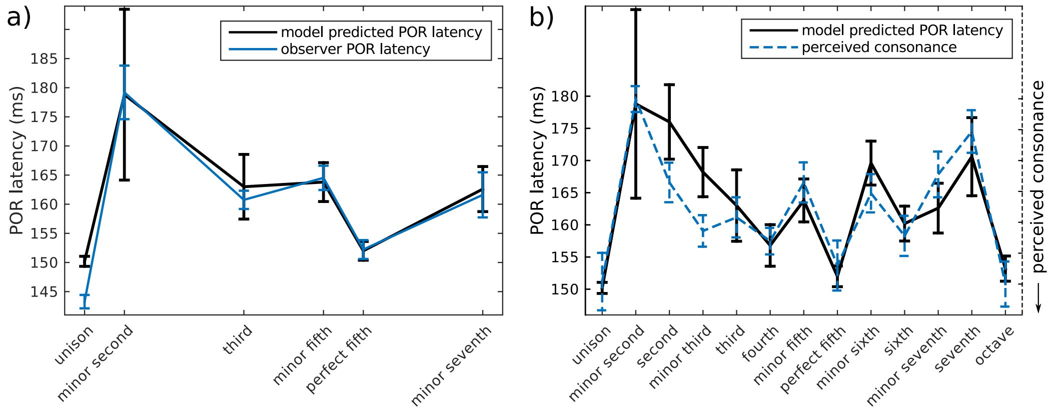


Figure 5.10: POR latency predictions for IRN dyads. a) Comparison between the predicted and observed latencies of the POR elicited by six IRN dyads. b) Extended prediction for the 13 dyads of the chromatic scale, in comparison with the consonance sensation typically reported for the same dyads. Stimuli consisted of the 13 IRN dyads with a fundamental frequency of $f_0 = 160\,\text{Hz}$ as described in §5.2.2. Predictions were averaged across five runs of the model, error bars of the predictions are standard deviations; notice the relatively large trial-to-trial variability of the minor second data, caused by the effect of cortical noise strengthen by the low signal-to-noise ratio of the second and third harmonic peaks of the SACF associated to that particular dyad (see Figure 5.5). Perceptual results were taken from [175], Fig. 2.

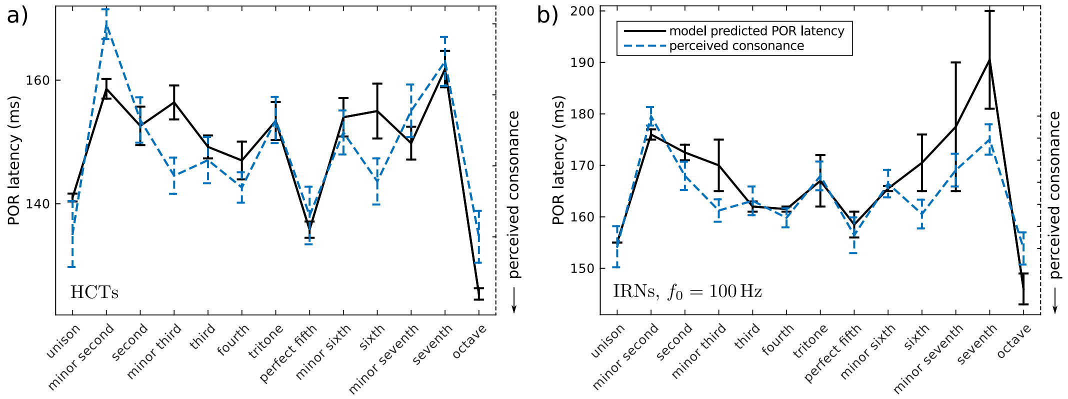


Figure 5.11: POR latency predictions for other dyads. Prediction for the 13 dyads of the chromatic scale, in comparison with the consonance sensation typically reported for IRN dyads. a) Dyads of harmonic complex tones; lower tone's fundamental frequency was set to $f_0 = 160\,\text{Hz}$. b) IRN dyads, with a lower tone's fundamental frequency $f_0 = 100\,\text{Hz}$. Perceptual results were taken from [175], Fig. 2.

Previous studies have linked the regularity of the SACF pattern elicited by dyads to their evoked consonance and dissonance percepts [176, 185, 196]. However, such studies did not provide for a mechanistic explanation of this relationship, and neither did they showed that harmonic collaboration has a direct effect on processing time.

Our model does not consider the influence of attention, cultural background, or musical training; it is designed to explain passive cortical pitch processing in the most possibly general fashion. Predictions on the POR latency are solely based on harmonic competition-collaboration in cortex and the decrease of the signal-to-noise ratio of the SACF due to the interaction between phase-locked activity in the auditory nerve characterising each tone. Thus, our model predicts that the correlation of consonance and dissonance with processing time may be largely a consequence of the pitch processing mechanisms in anterolateral portions of Heschl's gyrus, rather than a result of consonance preference or cultural exposure to consonant sounds.

On the other hand, our model does not perform any predictions over the cognitive preference for consonance widely reported in the literature. It could be argued that cognitive systems might show an innate preference for sounds that are processed faster, and thus require a smaller amount of resources, but we have seen that low-pitched tones are processed faster than high-pitched tones and humans do not show a particular preference for one or the other.

Thus, in the light of our modelling results, we propose that the emergence of the sensations of consonance and dissonance (but not the preference for one or the other) are universal, and that they are mainly sourced in pitch interactions during the generation of the SACF patterns and subsequent processing in alHG.

Conclusion

Vertical pitch interactions have several effects on the processing dynamics that can be regarded as non-linear, in the sense that the processing of a tone with two interacting pitch values cannot be simply expressed as the sum of the independent processing of the two tones.

The most outstanding of the non-linear effects is produced by vertical pitch interactions, resulting in the emergence of the sensations of consonance and dissonance [4, 172, 178]. A number of theories have been suggested during the last two centuries to account for this phenomenon. Early theories proposed that the sensation reflected the beating effect produced by the aggregation of sinusoidal waveforms with different frequencies [4, 172, 179, 191, 193, 194]; however, these theories were challenged by vast experimental evidence, suggesting that consonance is rather related to pitch processing [175, 185].

In agreement with these findings, our model introduces a new interpretation of consonance based on non-linear effects taking place at the pitch processing centre in auditory cortex. Vertical interactions essentially alter the time required for pitch perceptual decision making: dyads eliciting two simultaneous pitch values are processed slower than single-pitched tones [175] (see also §5.3.3.2). The exact difference in processing time is strongly correlated with the consonance/dissonance percept elicited by each tonal combination: consonant dyads are processed faster than dissonant dyads.

Previous studies have reported the correlation between the regularity of the SACF elicited by dyads to their evoked consonance and dissonance perception [176,181,182,185,196]. Here, we provide a mechanistic explanation of this effect. Harmonic facilitation during pitch processing occurs as a consequence of the interaction of the overlapping SACF patterns characterising consonant pitch combinations. Hence, only pitch processing models considering several peaks of the SACF representation are able to account for the effect (see §4.2.3.2).

Differences in pitch processing time might underlie the emergence of the sensations of consonance or dissonance (§5.3.3.2). Processing time is, however, only relevant during the onset of the decoding mechanisms in alHG; yet consonance is not an onset percept, but rather a continuous sensation present during the whole stimulation. Moreover, a mechanism measuring convergence time would be necessary to explain how processing time can be translated into an acoustic percept. Thus, further elements located at higher stages of the auditory processing hierarchy are necessary in order to explain how onset differences can give rise to a the continuous sensation characteristic of tint.

Although vertical interactions alter the signal-to-noise ratio of the harmonic series characterising each tonal component in the SACF representation (§5.1.6.3), they do not alter the final shape of the pitch rate cortical representation. Harmonic integration allows for the cortical system to successfully disentangle the non-linear interactions at the SACF and give rise to a linear cortical representation.

Although musical training enhances consonance and dissonance perception, and cultural background seems to be largely responsible for the association between consonance and pleasantness (§5.1.4), our experimental (§5.2.3) and theoretical (§5.3.4) results seem to indicate that non-linear effects are intrinsic to the cortical mechanisms of pitch processing, rather than the result of higher-level cognitive functions.

Chapter 6

Discussion

Conclusion

This thesis investigates the processing mechanisms underlying pitch perception in human auditory cortex at a mesoscopic scale. The main result is a theory describing cortical pitch processing as a competitive-cooperative process between balanced E/I ensembles, whose connectivity is specifically structured in a harmonic fashion. The theory introduces a cortical system that receives harmonic patterns of activation from subcortical areas and transforms them into a pitch rate representation, held in anterolateral Heschl's gyrus [27, 66, 68] and adjacent areas of planum temporale [53].

Our theory is embedded on a comprehensive model that presents some strong assumptions and rough idealisations of the underlying neuronal processes. However, it accounts for a wide range of psychoacoustical and electrophysiological data associated with pitch in a mechanistic fashion, for the first time to our current best knowledge. In addition, it introduced novel theoretical ideas linking the emergence of the sensations of consonance and dissonance with mechanisms underlying cortical pitch processing.

In this section we will summarise the most important conceptual results of our investigation, following the lines of the research questions formulated at the beginning of our investigation (see §1.2).

Neural representations of pitch along the multiple stages of the auditory pathway

A large part of this thesis has been dedicated to the identification of the different neural representations holding pitch-related information at different stages of the auditory pathway. Four relevant neural codes were identified during our investigation.

Temporal coding in the auditory nerve

Temporal coding, originated in the phase-locked responses of the auditory nerve to periodic stimuli (§2.1.1.2 and §3.1.3), may be perhaps the less controversial among the neural representations of pitch information often discussed in the literature. Evidence for the propagation of phase-locked activity to subcortical areas has been repeatedly reported in intracranial studies in mammals [15, 68] and in studies analysing brainstem responses in human subjects (§2.3.2.1).

It has been suggested [18] that spike trains with regular inter-spike-intervals could be used to perform comparisons between different pitch values, rendering it a plausible representation of pitch in cortex. Although phase-locking to low frequencies has been reported in human [16,88] and mammal [68] auditory cortex, there is no experimental evidence of cortical phase-locking above $\sim 200\,\text{Hz}$. Moreover, it seems difficult to argue how subjects with perfect pitch could use such code to perform absolute perceptual judgements (see §6.3.1).

Therefore, evidence seems to indicate that, although crucial during subcortical processing, phase-locked activity is not used as a representation of pitch in cortical areas (see also §3.2.3.4). Our theoretical ideas suggests that phase-locking is only relevant to convey pitch information prior to subcortical spectral analysis.

Spectro-temporal coding of the autocorrelation process

The SACF-like patterns described in §3.5.2 inevitably result from the inter-spike-interval analysis of the temporal representation described above, as originally established by the autocorrelations models [17, 123, 124], but also by more recent models of pitch [18]. Here, we suggest that such *spectro-temporal* representation holds pitch information at subcortico-cortical relays of the auditory pathway, and that it coexists with a pitch rate representation in auditory cortex.

Early versions of the autocorrelation model postulated the existence of *delay lines*, extensively criticised by their lack of a solid physiological basis [119] (see also §3.2.2.4). However, period-selectivity at neuronal complexes in cochlear nucleus and inferior colliculus have been shown to yield similar activation patterns in response to periodic spike trains in a physiologically sound manner [17, 127].

Experimental evidence for the presence of a spectro-temporal code in the mammal brain is relatively scarce, but spectral periodotopic arrangements have been found in the inferior colliculus of mammals (§2.1.2.3), and harmonic co-activation has been repeatedly reported in the mammal cortex [57, 62, 153] and in an fMRI study in humans [59, 61].

We showed earlier that the extra peaks at the lower harmonics grant a higher frequency resolution to the spectro-temporal representation (see §5.3.2.2). Moreover, the harmonic structure of the SACF might underlie the chromatic organisation of pitch [61, 119], and explain how spectral listeners perform their characteristic perceptual judgements (see §6.3.2).

Pitch-rate code

We use the term *pitch-rate code* to denominate the final pitch representation proposed in our model, where each neural block represents a single, determined pitch value. Pitch-selective neurons responding preferably to specific pitch values have been consistently reported in intracranial recording in the mammal brain [27, 66–68].

Our results suggest that a pitch rate code is held in anterolateral Heschl's gyrus, consistent with the position of equivalent dipoles active during pitch processing in MEG recordings [22, 91, 99] and with the location of pitch-selective neurons reported in fMRI recordings [19, 50, 54, 65].

There is little evidence for a consistent periodotopic organisation in human auditory cortex [46, 59, 61], suggesting that the pitch-rate code might not be topologically arranged in cortex (see §2.1.3.2). Although our model does present a fixed arrangement of the cortical populations conforming the pitch rate representation, model dynamics do not depend on the relative location of the neural ensembles; thus, our results are compatible with any specific topology.

Pitch rate coding is fundamental to explain the perception of subjects with perfect pitch [119], and might play an essential role in how absolute listeners judge tone intervals based on the pitch values of the two tones (see §6.3.2.)

Tonotopic code

Our model does not consider the tonotopic arrangements observed in Heschl's gyrus as an explicit representation of pitch in cortex. Pitch selective neurons are located only at the low-frequency edge of the tonotopic axis [19, 50, 54, 65] (see also §2.1.3.3) indicating that, although tonotopy plays an important role during the spectral analysis of the auditory nerve activity, it does not hold an explicit representation of pitch.

This hypothesis might be challenged when assessing the perception of pure tones. It has been suggested before that pure sinusoids do not share the same neural representation of pitch than stimuli with more complex spectral envelopes [149]. Although perceptual predictions of our model for pure tones are in line with psychoacoustic observations, pure tone dyads result in different consonance percepts than HCT or IRN dyads of harmonic complexes or iterated rippled noises [175,185,193]; an effect that cannot be explained by our model. A separate mechanism relying on tonotopy might underlie the perception of pure tones.

Neural mechanisms underlying cortical pitch processing in anterolateral Heschl's gyrus

This thesis introduced the hypothesis that neurons in anterolateral Heschl's gyrus are responsible for transforming the spectro-temporal code extracted in subcortical areas into a pitch-rate code. The neural mechanisms underlying this transformation conform a hierarchical structure consisting of two networks of neural ensembles, each characterised by a different effective time constant and presenting a different level of abstraction.

The first network, called *decoder*, reacts almost instantly to variations in the subcortical input and effectively computes the transformation between the two representations (§4.2.3.2). The second network, called *sustainer*, reacts only after the transformation process has converged and essentially modulates the decoder dynamics to avoid a continuous decoding when the subcortical input does not vary, presenting a higher inertia against changes than the lower network (§4.2.4.3). This architecture is stereotypical of systems of perceptual integration, where the decision propagates to higher-level systems with a larger inertia [150].

A similar structures has been introduced by previous studies modelling cortical integration of the autocorrelation output. Balaguer-Ballester and colleagues [125] introduced a cascade of two integrators performing successive integrations of the SACF with different time constants. In a later study [21], the authors introduced a top-down modulatory process that shaped the integration of the SACF according to the expectations of the incoming input (see also the details on the GPM model in §3.3.2).

The sustainer network uses the current extracted pitch value as expected pitch; a premise that was already present in Balaguer-Ballester's GPM [21]. However, whilst GPM considers the expectations as a pitch value, corresponding to the most likely outcome, the architecture of the new model encodes the likelihood of each pitch value separately along the columns of the sustainer, effectively allowing for complex expectation distributions encompassing several pitch values. This extension provides not only for the possibility of holding simultaneous pitch values in the cortical representation, but also allows for the introduction of complex priors from higher cognitive areas through the decrease of inhibitory drive in target columns at the sustainer (see §6.3.3).

The decoder network can be understood as a bank of interacting winner-takes-all integrators confronting harmonically related pitch values against each other (§4.4.2). The winner-takes-all architecture is a prototypical model describing the dynamics of systems in charge of perceptual decisions under ambiguous inputs ([150, 151]). Our system is an extension of those architectures allowing simultaneous pitch extraction processes to interact with each other. The transformation effectively performs a biologically sound *harmonic sieve* [119, 122], mapping harmonic patterns of activation into single-valued representations.

These neural mechanisms underlying harmonic inhibition are ultimately hardcoded in the connectivity strengths between excitatory and inhibitory ensembles in the decoder's network. These connectivity patterns could have been developed by early exposure to natural sounds during early stages of cortical formation [121]. Models using STDP have shown that harmonic connectivity patterns as the ones described in the model could naturally arise by a systematic exposure to harmonic patterns of activation [129], that might be systematically produced in subcortical areas in response to periodic sounds, widely present in nature.

Cortical pitch processing and the dynamics of the elicited auditory evoked fields

Collective activity in cortical ensembles can be indirectly measured through their elicited electromagnetic fields [24, 80]. Neural activity derived from our cortical model during pitch processing results in field predictions that quantitatively account for the equivalent dipole moments measured in anterolateral Heschl's gyrus reported by MEG studies (§4.3.2,§4.3.3). This predictive capability allows us to study in detail the origin of the field morphology, and the specific neural processing associated to several components of the auditory evoked field

Decoding dynamics and the POR

Decoder dynamics match the morphology of the pitch onset response (§4.2.7.2), describing this transient as the neuromagnetic fingerprint of cortical pitch extraction. The POR is typically elicited on pitch onset or pitch changes [22, 23, 91, 99, 105, 105], fully in line with the behaviour of the decoder, whose functioning is only triggered under changes in the subcortical input (§4.4.3.1 ,§4.4.3.4). The large depression characterising the POR morphology is a consequence of the aggregation of spectral evidence in the excitatory ensembles of the decoder (§4.2.3.2), gradually propagated from thalamus to cortex as a series of harmonically related peaks describing the pitch-related content of the auditory nerve activity.

After enough evidence in favour of a particular pitch value is present in the decoder, the decoding mechanisms is triggered by the action of inhibitory ensembles that effectively shunt cortical activity corresponding to harmonic peaks that do not correspond to the target pitch value (§4.2.3.2). This inhibitory action is responsible for the deflection of the POR transient, that peaks at the instant in which the inhibition begins to overcome the subcortical input (§4.2.3.3).

Thus, the mechanics of the model associate the POR peak latency with the convergence time of the decoding process: stimuli eliciting later PORs require more time for cortex to aggregate enough information before triggering the inhibitory action (§4.2.3.3). Our model sets the information limen to the resolution of the first three peaks of the harmonic series of the target period T up to a certain signal-to-noise ratio. The time required by the subcortical system to resolve the third peak of the harmonic series is $\sim 4T$, explaining the POR latency dependence with four times the perceived pitch typically reported in MEG experiments [22, 23, 91, 99].

Onset responses of other models of cortical pitch processing have been previously linked with the onset neuromagnetic fields. The derivative of the stabilised auditory image in AIM [128, 132, 135] has been related to the pitch onset response [108], although AIM did not succeed to perform quantitative predictions on the latency of the transient.

The POR has also been linked to the derivative of the activity at the integrator corresponding to the pitch value in the top-down modulated model [1, 21]. Although GPM did perform quantitative predictions on the latency and amplitude of the POR for IRNs [21] and ramped and damped sinusoids [1] (see also §3.4), these results are difficult to interpret.

First, the model associates the neuromagnetic responses to the activity at the population encoding the perceived pitch, whilst all cortical populations in the model should contribute similarly to the evoked fields [24, 80]. Moreover, Balaguer-Ballester's model does not provide for a mechanistic explanation of the source or behaviour of the transient. Our model extends previous results by explaining in detail the neural mechanisms underlying the rise of the POR.

Sustaining dynamics and the SF

The sustaining process exhibits similar dynamic properties as the sustained field, explaining the steady state of the neuromagnetic fields as the neuromagnetic fingerprint of pitch value short-term storing (§4.2.7.3). The sustained field is elicited during the last stage of pitch extraction, when the activity peaks corresponding to higher harmonics are being shunted (§4.2.4.3), explaining the paradoxical difference between the SF and the POR onset dynamics typically reported in MEG recordings [105, 107].

The sustainer network considers larger effective times of integration than the decoder network, due to the recurrent connections between inhibitory ensembles at each of the two cortical layers. This grants the sustainer a certain level or inertia, crucial to keep a steady pitch representation even in the presence of noise. Sustainer's recurrent connectivity provides for a mechanistic explanation for the offset delay observed in the pitch-related sustained field in MEG recordings [107].

The sustained field has been associated to the overall activity of the stabilised auditory image in AIM [108]. Although the SAI presents temporal properties explaining the offset delay of the SF [108, 132], the SAI onset is much faster than the onset of the sustainer, failing to explain the late rise of the response.

The role of the different regions of auditory cortex in pitch processing

Anatomical subdivisions of Heschl's gyrus

Despite evidence suggesting that the posteromedial section of Heschl's gyrus also plays an active role during pitch processing (see [200] for a review), MEG recordings locate the sources of the pitch-related onset and sustained responses in its anterolateral counterpart [22, 23, 105, 107, 108]. Our model, driven by MEG observations, suggests that cortical pitch processing is essentially carried out by two adjacent networks of neural ensembles located in alHG, corresponding to the cortical generators of the POR (§4.2.7.2) and the PSF (§4.2.7.2).

Intracranial recordings in human auditory cortex seem to indicate that the anterolateral section of HG is at a higher hierarchical level of the auditory pathway than the posteromedial section [26], and that only the later shows phase-locked responses [16, 55, 57, 58], suggesting that each anatomical subdivision might hold a different pitch representation [15, 16].

A plausible hypothesis is that pmHG acts as a relay between thalamus and the processing points in alHG, effectively low-passing the subcortical input before feeding it to the decoder (see A^{low} in §4.2.2.3). This would explain why phase-locking over 200 Hz vanishes only at this stage [16], why the tonotopic map is replicated twice in Heschl's gyrus [46, 54], and provide for a stable SACF-like spectral representation of pitch in cortex as described above (§6.1.1.2). More detailed experimental observations on the dynamics and structure of cortical activity in pmHG will be necessary before drawing more definitive conclusions.

Distributed processing and the pitch centre

One of the largest open controversies in current auditory neuroscience is the existence of a pitch centre: a neural conglomerate located in auditory cortex, generally extracting the perceived pitch elicited by different stimuli independently of their timbre or loudness [15, 27, 53, 57, 66, 68] (see also §2.1.3.3. Alternatively, cortex could extract pitch using a distributed processing network [26, 52], or there might be a collection of relays specialised in different types of spectral shapes, extracting pitch from stimuli with different timbres [15].

Our results show that there exists a biologically plausible mechanism able to extract a single pitch representation from a wide range of stimuli with different spectral contents, supporting the hypothesis of a general pitch centre in auditory cortex. However, the architecture of the pitch centre as introduced in our model is distributed in two adjacent cortical conglomerates (§4.2.1), and potentially includes regions beyond alHG holding lower level, timbre-dependent, representations of pitch, as detailed above in §6.1.4.1.

Cortical dynamics of pitch processing during consonance and dissonance perception

Subcortical representations of simultaneous pitch values

All three neural codes described in this work (see §6.1.1 above) are able to hold information characterising multiple simultaneous pitch values. In dyads, auditory nerve activity is phase-locked to the fundamental frequency of both tones (§5.1.3), and the derived SACF presents peaks of activation corresponding to the harmonic series of the two frequencies [174, 176, 181, 182].

Pitch interaction in the auditory nerve results in a decrease in the fidelity of the phase-locked responses when the fundamental frequencies of the tones in the dyad are not harmonically related [174, 176, 181, 182]. The effect further propagates to further neural representations, decreasing the signal-to-noise ratio of the associated SACF [174, 176]. The fidelity decrease is slightly stronger for dissonant than consonant tonal combinations, which results in subtle but significant variations in the hight of the peaks of the SACF representations associated to consonant and dissonant dyads [185, 196]. These differences are, however, strongly attenuated in the pitch-rate representation, where multiple pitch values can coexist without significant interferences (§5.3.2.2).

The emergence of the sensations of consonance and dissonance

Our results suggest that non-linear effects in multiple pitch extraction during cortical processing might be responsible for the emergence of the sensations of consonance and dissonance in dyads (§5.3.3.2). SACF-like spectral representations associated to consonant tone combinations show partially overlapping harmonic structures, that facilitate the extraction of the participating pitch values [185, 196]. Dissonant combinations result in uncorrelated harmonic series that are independently extracted, resulting in a slower decoding onset than their consonant counterparts [185, 196]. Although our cortical theory does not explicitly introduce a general representation of consonance in cortex, we hypothesise here that these sensations might be a subproduct of such time processing differences (§5.3.4).

Previous models on consonance perception have used the regularity of the SACF [196] or the signal-to-noise ratio of the represented harmonic series [176] to predict perceived consonance. However, these models do not provide for a mechanistic explanation of this relationship.

Although the association between pleasantness and consonance is not universal, this association is extraordinary common to many cultures [172,180,186,187]. A possible explanation for the spread of consonance preference might be the wide dissemination of Indoeuropean languages, whose phonetic content primes consonant tonal combinations over their dissonant counterparts [180]. Alternatively, the human brain might show a biased, but not decisive, preference towards periodic SACF-like representation, as a result to the continuous exposure to single-pitched tones present in the nature. This preference might be established at later stages of cortical development, explaining why infants do not show a particular preference towards consonant combinations [189].

Limitations of the cortical model and future extensions

The main contribution of this thesis is the introduction of a novel model of cortical pitch processing based on the integration across several harmonics of a subcortical SACF-like spectral representation. This section summarises the range of applicability of the model and its limitations, introducing further extensions that will be developed on future work.

Limitations introduced by the computational implementation

The dynamics of the cortical model are too complex to be analytically tractable; thus, all model predictions were obtained through numerical simulations that introduced several limitations in the range of applicability of the model. Those limitations are, however, not intrinsic to the proposed mechanisms, but rather a consequence of this particular computational implementation.

Loss of inhibitory accuracy in the high-frequency range

An important limitation is introduced by the discretisation of the space of the periods in the SACF and the pitch representations of the cortical model. The discretisation is responsible for a noticeable loss of accuracy, specially pronounced in some stimuli, in the resolution of pitch values within the high-frequency range $f_0 \gtrsim 500\,\mathrm{Hz}$ (see Figure 4.7 and §4.3.1.1).

The implementation of the model used in this thesis considers a discretisation of the period space in $N = 250$ different segments, ranging from $\delta t_0 = 0.5\,\mathrm{ms}$ to $\delta t_N = 33\,\mathrm{ms}$. This configurations results in a maximum standard error of $\Delta = 0.065\,\mathrm{ms}$, that scales up to a 4 ms error in the inhibition of the farther lower harmonics (see §4.3.1.1 for details). This systematic error can be reduced by considering a lower range of periods or a larger number of chunks in the period space; current parameters were chosen as a trade off between resolution, range of applicability, and computational time complexity.

Since discretisation is necessary to construct a computational implementation of the model, further developments should address this issue by increasing the resolution of the discretisation so that the expected derived errors lie over the actual limits of pitch measured in psychoacoustic experiments [5,41,201].

Lower limit of pitch set to 100 Hz

A second limitation is introduced by the lower bound in the period space $\delta t_n \geq 33\,\mathrm{ms}$. The cortical model integrates over three harmonic peaks of the SACF representation in order to perform a perceptual decision. Our implementation is thus only able to fully account for the integratory process of pitch values corresponding to $T > 11\,\mathrm{ms}$.

Stable perceptual predictions for most stimulus types can be robustly extracted using only two harmonic peaks, for periods up to $T \sim 16\,\mathrm{ms}$; however, the dynamics associated with the extraction process in those cases are not fully reliable. This lower limit of applicability could be easily extended to the actual lower limit of pitch by considering periods up to $\delta t_N = 10\,\text{ms}$ [77], at the cost of a large increase in the number of variables of the model or a decrease in the resolution of the period space.

Limitations introduced by the formulation of the subcortical processing

The across-stimulus variability of the SACF peak amplitude has a noticeable effect on the dynamics of the cortical model, that would not be expected in a more idealised representation of the input. Amplitude variations are vastly reduced through the regularisation procedure, but perceptual results indicate that a more adaptive regularisation procedure might be necessary in order to extend our results to a larger range of stimuli, as detailed next.

Baseline removal

Despite the wide success of the model to explain perceptual data in single and multiple pitch values, the perception of alternated-phase HCTs (§4.3.1.3) requires an adaptive, stimulus-dependent, SACF baseline removal parameters b_0 (see §4.2.2.3). Adaptive baseline removal could be implemented through global inhibition triggered by the total activity at the SACF [148, 152]; however, most stimuli types (pure tones, HCTs, click train, filtered IRNs, and even dyads) yield satisfactory perceptual results with the same baseline $b_0 = 0.35$ (§4.3.1), despite the fact that they present very different overall regularised SACF activities (see Figure 4.3).

A more sophisticated mechanism, probably involving a top-down control from the decoder network, might hence underlie SACF baseline removal. Future work should address this issue introducing cortico-thalamic efferents regulating subcortical processing according to the signal-to-noise ratio arriving in the cortical network. Baseline removal might be performed globally, over the whole SACF representation, or selectively, operating only on key cochlear channels.

Rescaling gain

Similarly, an adaptive SACF rescaling gain A_0 seems necessary in order to explain the compatibility of the cortical model with both, single tones and dyads (see 5.3.1). We suggest that this might be mediated by an intermediate mechanism, integrating the regularised SACF in order to increase the signal to noise ratio of the representation until the activity values are large enough as to trigger a significant response in the decoder network; explaining as well why a systematic delay of 25 ms was found between the latency predictions associated with dyads and single tones.

However, long integration time constant at the intermediate level would frustrate the responsiveness of the cortical model to input changes. Future work should address this problem by the use of adaptive integration time constants regulated by the decoder layer, following the formulation of the top-down modulated model [21]. The use of adaptive integration constants might also extend the psychoacoustic results of the current formulation to the wide set of stimuli explained by the top-down model [21]. How this adaptive behaviour could be implemented in a biophysically sound manner is, however, still unclear.

Future work should also evaluate the impact of using the subcortical slope coincidence detectors by Huang and Rinzel [18] prior to the spectral analysis. Huang's model has been shown to reduce the impact of timbre and loudness on the regular spike trains of the time code at the auditory nerve, potentially facilitating the regularisation process.

Limitations intrinsic to the formulation of the model

Frequency resolution in dyads

The minimum separation between the two pitch values comprising a dyad required for the cortical model to independently resolve both tones (see Figure 5.7B) is much restrictive than the limits described by psychophysical data [4]. This limitation is a consequence of the architecture of the decoder rather than a side effect of discretisation.

The decoder is designed to integrate activity across the first three peaks of the SACF harmonic series representing the pitch of the stimulus (§4.2.3.2). However, these peaks are sometimes too broad to be independently resolved when two overlapping harmonic series are represented in the SACF.

The width of the peaks does not depend on the discretisation of the period space, but on the spectral properties of the auditory nerve activity [124]; compare, for instance, the top-left panels of Figures 3.2A and 3.2C, corresponding to the SACF harmonic peaks associated with pure tones and IRNs, respectively. The peaks associated with pure tones are much thicker than the peaks associated with IRNs, although the resolution of the period space is the same in both cases.

When the first peaks of two harmonic series overlap in the SACF representation, aggregating information across the second and third peaks allows the cortical model to independently resolve the pitch values corresponding to each of the series (§5.3.2.2). This is, however, not possible when the frequency separation between the two pitch values is so small than the second and third peaks of the series overlap in the SACF representation.

This limitation could be addressed by increasing the number of considered harmonic peaks during cortical integration. This strategy would, however, require processing times that might scale up to six or seven harmonics, much longer than the time constants observed in MEG recordings [22, 23].

Future research should address this issue by using a generative approach [29, 168] to regulate the number of harmonic peaks taken into consideration during processing, in resemblance with GPM, which also used a generative approach to regulate the size of the temporal window of integration during pitch processing [21].

A generative version of our model could use the predictions performed using the first three harmonics of the series as a first approximation of the extracted pitch. SACF harmonic series corresponding to the provisional pitch could then be generated downwards and compared against the actual subcortical input. This comparison is already performed by the inhibitory ensembles in the decoder by means of selective inhibition (§4.2.3.2). Prediction error could then be used to refine the extracted pitch using the non-inhibited peaks at the excitatory populations at the decoder.

The behaviour of the generative system should be modelled according to psychophysical data detailing: 1) the actual minimum pitch distance required to independently perceive the two tones comprising a dyad, as a function of the fundamental frequency of the lower tone; and 2) the minimum number of period cycles necessary for a listener to robustly resolve the pitch of the two tones in each dyad. This data could be acquired in a perceptual experiment presenting dyads with different frequency combinations to listeners that must report whether they perceive a single pitch percept or two concurrent tones.

Cortical representation of pitch strength

Decoding dynamics introduced in our model seem to be only weakly affected by pitch strength (see Figure 4.12B), and the cortical pitch-rate representation is widely unaffected by this property. However, the effect of pitch strength on the POR amplitude [1, 22] and the PSF depth [1, 108] seems to indicate that this sensation might be sourced in the same processing centres as the pitch value.

The indifference of our model to pitch strength is mostly due to this quality being is correlated to the spectral contour (rather than to the fundamental frequency) of the stimulus waveform. Pitch strength variations are not reflected in the characteristic periods of the peaks of the SACF fall but in the relative height of the SACF harmonic series, which do not affect the mechanisms of pitch extraction described in our model.

Pitch strenght information is still present in the subcortical representation after the regularisation process, which preserves the overall structure of the SACF and only affects its absolute and the average activity. Future work should address this effect and determine if the cortical mechanisms described in this thesis can be extended to hold a representation of pitch strength; or if, on the contrary, pitch strength is reflected in a distinct representation (i.e, in the spectro-temporal code, see §6.1.1.2).

Open problems and open theories

Despite the relative success of our model in explaining many perceptual and electrophysiological results in pitch perception, there are still a large number of open problems, generally concerned with inter-subject differences or the perception of higher cognitive auditory objects, that were not addressed in our investigation. This section explores, in a speculative and qualitative manner, how some of these problems might be tackled in future work.

Absolute pitch

Although the typical listener is able to perform judgements about the relative pitch value of two or more tones, few human subjects are able to label the pitch of a sound without using a reference. Such rare ability is commonly called perfect or absolute pitch [78]. The neural substrate of perfect pitch is one of the most intriguing open questions in auditory neuroscience [5]. Here, we attempt to provide a potential heuristic explanation of this phenomenon in terms of neural connectivity and consciousness.

The integrated information theory of consciousness (ITT), first introduced by Tononi [202], postulates that conscious systems are characterised by their integrated information; i.e, the amount of information that would be lost when subdividing the system in smaller unconnected subsystems. ITT determines the inclusion of a unit connected to a system as part of the conscious system depending on the connectivity pattern linking the two bodies. Strong and complex connections usually yield inclusion, whilst units linked by simple feed-forward-like connections do not form part of the conscious body (see Figure 6.1A).

Relative pitch judgements could be performed by a neural system comparing the pitch representation of the prove and reference tone [203] (see also Figure 6.2A). In the ITT framework, relative pitch listeners' conscious network would include the output of the system comparing the pitch representations of the two tones, but it would not have access to the units holding the pitch representation themselves (see Figure 6.1B). On the contrary, subjects with perfect pitch would present an integrated architecture, where the units at the sustainer, holding the pitch-rate representation, would be part of the conscious body (see Figure 6.1C).

Accordingly, subjects with absolute pitch show higher white matter connectivity and track volume in posterior superior and middle temporal gyri, with a specially prominent hyperconnectivity in the left hemisphere [204], and an increased volume and activity levels during pitch processing in the left Heschl's gyrus [205].

Spectral and fundamental pitch listeners

The concepts of *spectral* (f_{SP}) and *fundamental listeners* (f_0) were introduced earlier in §2.2.3.5. The difference between both types of subjects is explicit when they evaluate the relative pitch values of two HCTs with missing fundamentals: f_{SP} listeners judge these intervals according to spectral clues (i.e, the the lower harmonic present in each tone), whilst f_0 listeners judge these intervals according to the actual pitch value of the tones (i.e, the fundamental frequency of the HCTs) [4, 45]. Their evaluations differ when the HCT with the lower fundamental frequency has a higher lower harmonic than the HCT with the higher fundamental frequency (see Figure 2.6A).

Most listeners do not show a consistent f_{SP}/f_0 perceptual mode, but rather a preferred tendency towards one or the other that can be measured in a continuous scale ranging from absolute f_{SP} to absolute f_0. Our cortical model suggests that two different representations of pitch coexists in cortex: a spectral representation, that could be used to identify the lowest harmonic of the HCTs as long as they are independently resolved in the cochlea (see Figure 4.3C for an example), and a pitch-rate representation, that can be used to perform f_0-like perceptual judgements.

Relative pitch evaluations could be modelled as the resolution of a decision system receiving weighted inputs from two subsystems judging pitch differences according to each of the two representations available in cortex (see Figure 6.2). A biophysically plausible neural system performing interval judgements using a pitch-rate representation was introduced recently by Huang and colleagues [203].

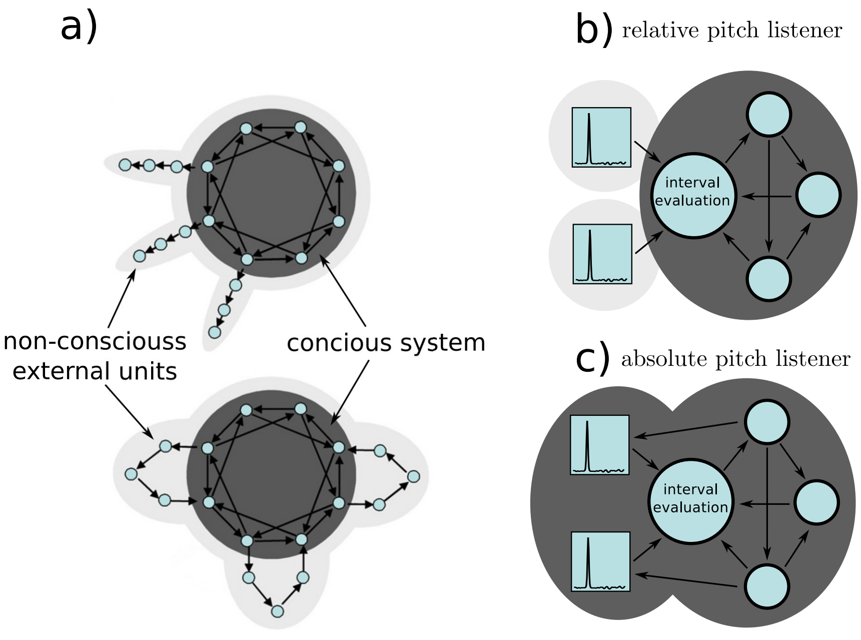


Figure 6.1: ITT and schematics for absolute pitch. a) Two system decomposed into a conscious entity and external units not being part of the conscious system. b) Schematic architecture associated to a candidate relative pitch listener; feed-forward connections between units holding the pitch-rate representation and the interval judgement system do not convey enough integrated information as to include the units in the conscious system. c) Schematic architecture associated to a candidate absolute pitch listener; increased connectivity complexity make the units be part of the conscious body. Panel A adapted from [202], Fig 4.

In this framework, f_{SP}/f_0 preference could be modelled as the relative weight of each of the two perceptual subsystems: an absolute f_0 listener would rely only in the judgements taken by the subsystem using the pitch-rate representation, whilst absolute f_{SP} listeners would rely only in the spectral representations.

Spectral/fundamental mode tendency is strongly correlated with the relative volume between the hemispheric replications of lateral Heschl's gyrus: f_0 listeners show a larger left than right lHG, whilst f_{SP} show an increased right lHG [45]. This correlation might be a result of a hemispheric specialisation in each of the two representations, pitch-rate being more prominent in the left hemisphere, and spectral representations being more prominent in the right hemisphere; explaining the perceptual differences and their anatomical correlations. Accordingly, subjects with absolute pitch (see above) enhanced left Heschl's gyrus [205].

Focused attention

Most sensory processing is carried out in a passive fashion. In order to reproduce the conditions of such passive processing, typical MEG experiments (e.g. [1, 22, 99, 108]) are designed to deviate the attention of the subjects from the auditory stimuli during the measurements. Accordingly, our model was devised to consider automatic, passive, cortical pitch processing.

However, focused attention plays a crucial role in sensory processing, and most psychoacoustic experiments are performed on attending subjects [5]. Although a detailed description of the cortical dynamics of pitch processing in attentional states is out of the scope of this thesis, here we will briefly describe in a heuristic manner how top-down attentional mechanisms might be incorporated to our cortical model in future work.

Focused attention is crucial to filter out selective inputs from a noisy, distracting environment [206], and can modify pitch perception in several ways. For instance, attending subjects are able to independently resolve up to five different harmonics of a HCT [4, 5].

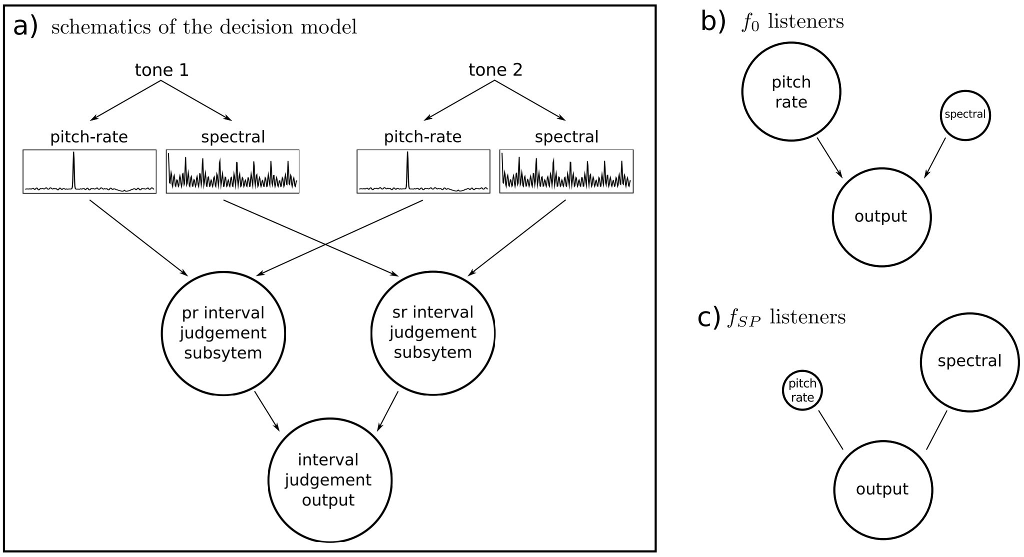


Figure 6.2: Hypothetical architecture of a dual system for interval judgements. a) Architecture of the system: subsystem receive inputs from a pitch-rate/spectral representation of the two consecutive notes. The integrator at the bottom aggregates the decision of the two systems to perform the final perceptual decision. b) Suggested architecture characterising fundamental listeners: the subsystem using the pitch-rate representation has a larger weight in the decision process. c) Suggested architecture characterising spectral listeners.

Moreover, focused attention can speed up auditory processing and, when the attention is selective to certain frequency ranges, reduce the detection threshold for sounds in the target frequencies [207]. Similarly, attention has been shown to affect the N100 latency [208], to mildly modulate the N100 amplitude [98], and to severely increase the sustained field depth [98].

Focused attention is conveyed by top-down cortical efferents sourced in higher-level cognitive areas and targeting cortical and subcortical systems [206, 209]. Future work could attempt to describe top-down attention in the cortico-cortical modality as localised selectively inhibition towards target ensembles in the sustainer network (see §4.4.4).

For instance, the resolution of multiple peaks of a HCT could be achieved by selectively inhibiting inhibitory neurons at the sustainer characterising the frequency range of the harmonics to be resolved, allowing multiple harmonically related peaks of activation to temporarily coexist in the cortical representation. A similar mechanisms would also allow to decrease the detection thresholds in certain frequency domains, effectively blocking potential distractors in noisy or acoustically-crowded environments.

An attention mechanism based on selective inhibition towards sustainer's inhibitory ensembles could also explain the increase in the sustained field observed in attentive states [98].

Closing remarks

Results shown in this thesis suggest that the pitch-related neuromagnetic fields observed in the anterolateral section of Heschl's gyrus during pitch processing might reflect pitch extraction from a spectral representation of pitch-related information similar to that of the autocorrelation models. During our investigation, we have developed a cortical model to implement such a pitch extraction procedure, and showed that the derived mechanisms reproduce the dynamics of the pitch onset response and the pitch related sustained field.

Our model fills the gap between the pitch-rate representation typically reported in intracranial cortical studies (e.g, [27, 66–68]) and the spectral representation characteristic of autocorrelation models (e.g. [17, 124]) and other theoretical models of pitch processing [18]. Previous studies addressing this transformation in cortex typically focus on isolating the first peak of the SACF whilst avoiding the read-out of the peak corresponding to the self-correlation of the stimulus [10, 125].

These and other more biologically based models have successfully explained the perception of a wide set of stimuli, and mirror several aspects of neuromagnetic data [1, 21]. However they do not provide an detailed enough mechanistic explanation in terms of realistic networks of the emergence and behaviour of the auditory evoked fields.

Our model explains the emergence of the pitch onset response as the cortical accumulation of evidence towards a determined perceptual decision, followed by a filtering of the non-necessary information after pitch extraction (§4.2.7.2). This link provides a theoretical association of the latency of the pitch onset response with processing time.

An additional limitation of models focused on the enhancement of the first peak of the autocorrelation function is that they are unable to account for the perception of simultaneous pitch values [119]. Harmonic sieve strategies address this issue by mapping harmonic series to their corresponding pitch values, allowing for the extraction of multiple pitches from a single SACF representation [119, 122]. Cortical mechanisms at the decoder can be described as a biologically sound and detailed implementation of a harmonic sieve process.

Our implementation predicts several non-linear effects resulting from the interaction of the decoded tonal components. Specifically, we showed that processing time is dramatically affected by the frequency relationship between the two pitch values, and that processing time was strongly correlated with the reported consonance and dissonance of each tone combination (§5.3.3.2). This is in full agreement with experimental observations on the dependence of the POR latency with consonance [175] (§5.2.3).

Despite the relatively large success of the model to explain perceptual effects as the emergence of consonance and dissonance, and neuromagnetic results as the dependence of the POR's latency with pitch, our formulation is still far from capturing the behaviour of cortical pitch processing in a comprehensive manner. Stimuli presenting more complex temporal structure require more sophisticated mechanisms that are not yet present in our model. More importantly, our model is unable to account for context-dependent effects on pitch processing [41], or how pitch processing affects the formation of higher order auditory objects [210]. Further modelling developments should attempt to explain how adaptive and context-specific behaviours arise in auditory cortex, rendering the vast richness of the auditory experience.

Bibliography

[1] A. Tabas, A. Siebert, S. Supek, D. Pressnitzer, E. Balaguer-Ballester, and A. Rupp, "Insights on the Neuromagnetic Representation of Temporal Asymmetry in Human Auditory Cortex," *PLoS One*, vol. 11, no. 4, p. e0153947, 2016.

[2] A. Tabas, A. Rupp, and E. Balaguer-Ballester, "Competition Between Cortical Ensembles Explains Pitch-Related Dynamics of Auditory Evoked Fields," in *International Conference on Artificial Neural Networks*, pp. 9886:314–321, 2016.

[3] B. Kolb and I. Q. Whishaw, *An Introduction to Brain and Behavior*. Worth Publishers, third edit ed., 2005.

[4] H. L. F. Helmholtz and A. J. Ellis, *On the Sensations of Tone as a Physiological Basis for the Theory of Music*. Dover Publications, 2009.

[5] B. C. J. Moore, *An Introduction to the Psychology of Hearing*. Academic Press, fifth edit ed., 2003.

[6] J. Schnupp, I. Nelken, and A. King, *Auditory Neuroscience: Making sense of sound*. MIT Press, first edit ed., 2011.

[7] A. A. Ludwig, M. Fuchs, E. Kruse, B. Uhlig, S. A. Kotz, and R. Rübsamen, "Auditory Processing Disorders with and without Central Auditory Discrimination Deficits," *Journal of the Association for Research in Otolaryngology*, vol. 15, no. 3, pp. 441–464, 2014.

[8] J. C. Stévens and J. W. Hall, "Brightness and loudness as functions of stimulus duration," *Perception & Psychophysics*, vol. 1, no. 9, pp. 319–327, 1966.

[9] R. D. Patterson, E. Gaundrain, and T. C. Walters, "The Perception of Family and Register in Musical Tones," in *Music Perception* (M. Riess Jones, R. R. Fay, and A. N. Popper, eds.), vol. 36 of *Springer Handbook of Auditory Research*, pp. 13–50, New York, NY: Springer New York, 2010.

[10] A. de Cheveigné, "Pitch perception models - a historical review," in *International Conference of Acoustics*, vol. 71, pp. 1–69, 2004.

[11] W. A. Yost, "Pitch of iterated rippled noise," *Journal of the Acoustical Society of America*, vol. 100, no. 1, pp. 511–518, 1996.

[12] W. A. Yost, "Pitch strength of iterated rippled noise," *Journal of the Acoustical Society of America*, vol. 100, no. 1, pp. 3329–3335, 1996.

[13] P. Heil and A. J. Peterson, "Spike timing in auditory-nerve fibers during spontaneous activity and phase locking," *Synapse*, 2016.

[14] M. S. Malmierca, T. A. Hackett, B. R. Schofield, R. A. Altschuler, and S. E. Shore, "Structural and functional organization of the auditory brain," in *The Oxford Handbook of Auditory Science: The Auditory Brain* (A. R. Palmer and A. Rees, eds.), pp. 9—-89, Oxford University Press, 2010.

[15] K. M. M. Walker, J. K. Bizley, A. J. King, and J. W. H. Schnupp, "Cortical encoding of pitch: Recent results and open questions," *Hearing Research*, vol. 271, no. 1-2, pp. 74–87, 2011.

[16] J. F. Brugge, K. V. Nourski, H. Oya, R. a. Reale, H. Kawasaki, M. Steinschneider, and M. a. Howard, "Coding of repetitive transients by auditory cortex on Heschl's gyrus.," *Journal of neurophysiology*, vol. 102, pp. 2358–74, oct 2009.

[17] R. Meddis and L. P. O'Mard, "Virtual pitch in a computational physiological model," *The Journal of the Acoustical Society of America*, vol. 120, no. 6, p. 3861, 2006.

[18] C. Huang and J. Rinzel, "A Neuronal Network Model for Pitch Selectivity and Representation," *Frontiers in Computational Neuroscience*, vol. 10, no. June, pp. 1–17, 2016.

[19] H. Penagos, J. R. Melcher, and A. J. Oxenham, "A neural representation of pitch salience in nonprimary human auditory cortex revealed with functional magnetic resonance imaging.," *The Journal of Neuroscience*, vol. 24, pp. 6810–5, jul 2004.

[20] R. Patterson, "Auditory images: How complex sounds are represented in the auditory system," *The Journal of the Acoustical Society of Japan*, vol. 21, no. 4, pp. 183–190, 2000.

[21] E. Balaguer-Ballester, N. R. Clark, M. Coath, K. Krumbholz, and S. L. Denham, "Understanding pitch perception as a hierarchical process with top-down modulation," *PLoS Computational Biology*, vol. 5, p. e1000301, mar 2009.

[22] K. Krumbholz, R. D. Patterson, A. Seither-Preisler, C. Lammertmann, and B. Lütkenhöner, "Neuromagnetic evidence for a pitch processing center in Heschl's gyrus," *Cerebral Cortex*, vol. 13, no. 7, pp. 765–772, 2003.

[23] S. Ritter, H. Günter Dosch, H.-J. Specht, and A. Rupp, "Neuromagnetic responses reflect the temporal pitch change of regular interval sounds.," *NeuroImage*, vol. 27, pp. 533–43, sep 2005.

[24] M. Hämäläinen, R. Hari, and R. Ilmoniemi, "Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the working human brain," *Reviews of Modern Physics*, vol. 65, no. 2, pp. 414–497, 1993.

[25] H. Lindén, T. Tetzlaff, T. C. Potjans, K. H. Pettersen, S. Grün, M. Diesmann, and G. T. Einevoll, "Modeling the spatial reach of the LFP," *Neuron*, vol. 72, pp. 859–872, dec 2011.

[26] S. Kumar, W. Sedley, K. V. Nourski, H. Kawasaki, H. Oya, R. D. Patterson, M. A. H. III, K. J. Friston, and T. D. Griffiths, "Predictive Coding and Pitch Processing in the Auditory Cortex.," *Joural of Cognitive Neuroscience*, vol. 23, no. 10, pp. 3084–3094, 2011.

[27] J. K. Bizley, K. M. M. Walker, A. J. King, and J. W. H. Schnupp, "Neural ensemble codes for stimulus periodicity in auditory cortex.," *The Journal of Neuroscience*, vol. 30, no. 14, pp. 5078–5091, 2010.

[28] P. Dayan and L. Abbott, *Theoretical Neuroscience: Computational And Mathematical Modeling of Neural Systems.* Computational Neuroscience, Massachusetts Institute of Technology Press, sixth edit ed., 2005.

[29] K. Friston, "Learning and inference in the brain.," *Neural Networks*, vol. 16, pp. 1325–52, nov 2003.

[30] W. Gerstner, W. M. Kistler, R. Naud, and L. Paninski, *Neuronal Dynamics.* Cambridge University Press, 1st ed., 2014.

[31] S. Hochstein and M. Ahissar, "View from the Top: Hierarchies and Reverse Hierarchies Review," *Neuron*, vol. 36, no. 5, pp. 791–804, 2002.

[32] P. Belin, M. Zilbovicius, S. Crozier, L. Thivard, A. Fontaine, M.-C. Masure, and Y. Samson, "Lateralization of speech and auditory temporal processing," *Journal of Cognitive Neuroscience*, vol. 10, no. 4, pp. 536–540, 1998.

[33] D. Poeppel, "The analysis of speech in different temporal integration windows: Cerebral lateralization as 'asymmetric sampling in time'," *Speech Communication*, vol. 41, no. 1, pp. 245–255, 2003.

[34] R. J. Zatorre and P. Belin, "Spectral and temporal processing in human auditory cortex.," *Cerebral Cortex*, vol. 11, pp. 946–53, oct 2001.

[35] R. J. Zatorre, P. Belin, and V. B. Penhune, "Structure and function of auditory cortex: music and speech.," *Trends in Cognitive Sciences*, vol. 6, pp. 37–46, jan 2002.

[36] R. Meddis and E. a. Lopez-poveda, "Computational Models of the Auditory System," in *Computational Models of the Auditory System* (R. Meddis, E. A. Lopez-Poveda, R. R. Fay, and A. N. Popper, eds.), vol. 35, ch. 2, pp. 7–39, Springer Science+Business Media, 2010.

[37] L. Chittka and A. Brockmann, "Perception space - The final frontier," *PLoS Biology*, vol. 3, no. 4, pp. 0564–0568, 2005.

[38] W. P. Shofner and G. Selas, "Pitch strength and Stevens's power law," *Perception & Psychophysics*, vol. 64, pp. 437–450, apr 2002.

[39] M. S. a. Zilany, I. C. Bruce, and L. H. Carney, "Updated parameters and expanded simulation options for a model of the auditory periphery," *The Journal of the Acoustical Society of America*, vol. 135, no. 1, pp. 283–286, 2014.

[40] A. J. Oxenham, "Pitch perception," *The Journal of Neuroscience*, vol. 32, pp. 13335–8, sep 2012.

[41] A. J. Oxenham, C. Micheyl, M. V. Keebler, A. Loper, and S. Santurette, "Pitch perception beyond the traditional existence region of pitch.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, pp. 7629–34, may 2011.

[42] J. Burkard, R., Don, M., Eggermont, *Auditory Evoked Potentials: Basic Principles and Clinical Application.* Point (Lippincott Williams and Wilkins) Series, Lippincott Williams & Wilkins, 2007.

[43] A. de Cheveigné, "Pitch Perception," in *Oxford Handbook of Auditory Science: Hearing* (C. J. Plack, ed.), pp. 71–104, Oxford University Press, 2010.

[44] L. Cedolin and B. Delgutte, “Spatiotemporal Representation of the Pitch of Harmonic Complex Tones in the Auditory Nerve,” *The Journal of Neuroscience*, vol. 30, pp. 12712–12724, sep 2010.

[45] P. Schneider, V. Sluming, N. Roberts, M. Scherg, R. Goebel, H. J. Specht, H. G. Dosch, S. Bleeck, C. Stippich, and A. Rupp, “Structural and functional asymmetry of lateral Heschl’s gyrus reflects pitch perception preference.,” *Nature Neuroscience*, vol. 8, pp. 1241–7, sep 2005.

[46] M. Saenz and D. R. M. Langers, “Tonotopic mapping of human auditory cortex,” *Hearing Research*, vol. 307, pp. 42–52, 2014.

[47] G. D. Pollak, J. X. Gittelman, N. Li, and R. Xie, “Inhibitory projections from the ventral nucleus of the lateral lemniscus and superior paraolivary nucleus create directional selectivity of frequency modulations in the inferior colliculus: a comparison of bats with other mammals.,” *Hearing Research*, vol. 273, pp. 134–44, mar 2011.

[48] J. A. Winer, “The human medial geniculate body,” *Hearing Research*, vol. 15, no. 3, pp. 225–280, 1984.

[49] R. B. Buxton, *Introduction to Functional Magnetic Resonance Imaging: Principles and Techniques*. Cambridge University Press, 2002.

[50] T. D. Griffiths, S. Uppenkamp, I. Johnsrude, O. Josephs, and R. D. Patterson, “Encoding of the temporal regularity of sound in the human brainstem.,” *Nature Neuroscience*, vol. 4, pp. 633–7, jun 2001.

[51] M. Moerel, F. De Martino, K. Uğurbil, E. Yacoub, and E. Formisano, “Processing of frequency and location in human subcortical auditory structures.,” *Scientific reports*, vol. 5, p. 17048, 2015.

[52] T. A. Hackett and J. H. Kaas, “Auditory Cortex in Primates: Functional Subdivisions and Processing Streams,” in *The Cognitive Neurosciences III*, pp. 215–232, MIT Press, 3rd editio ed., 2004.

[53] D. Bendor, “Does a pitch center exist in auditory cortex?,” *Journal of Neurophysiology*, vol. 107, pp. 743–6, feb 2012.

[54] M. Moerel, F. De Martino, and E. Formisano, “Processing of natural sounds in human auditory cortex: tonotopy, spectral tuning, and relation to voice sensitivity.,” *The Journal of Neuroscience*, vol. 32, pp. 14205–16, oct 2012.

[55] M. Steinschneider, D. H. Reser, Y. I. Fishman, C. E. Schroeder, and J. C. Arezzo, “Click train encoding in primary auditory cortex of the awake monkey: evidence for two mechanisms subserving pitch perception.,” *The Journal of the Acoustical Society of America*, vol. 104, pp. 2935–2955, dec 1998.

[56] D. Bendor and X. Wang, “Neural Coding of Periodicity in Marmoset Auditory Cortex,” *Journal of Neurophysiology*, vol. 103, no. 4, pp. 1809–1822, 2010.

[57] X. Wang and K. M. M. Walker, “Neural Mechanisms for the Abstraction and Use of Pitch Information in Auditory Cortex,” *Journal of Neuroscience*, vol. 32, no. 39, pp. 13339–13342, 2012.

[58] C. A. Atencio and C. E. Schreiner, “Auditory cortical local subnetworks are characterized by sharply synchronous activity.,” *The Journal of Neuroscience*, vol. 33, no. 47, pp. 18503–14, 2013.

[59] M. Moerel, F. De Martino, R. Santoro, K. Ugurbil, R. Goebel, E. Yacoub, and E. Formisano, "Processing of natural sounds: characterization of multipeak spectral tuning in human auditory cortex.," *The Journal of Neuroscience*, vol. 33, pp. 11888–98, jul 2013.

[60] T. M. Shackleton and R. P. Carlyon, "The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination.," *The Journal of the Acoustical Society of America*, vol. 95, no. 6, pp. 3529–3540, 1994.

[61] M. Moerel, F. De Martino, R. Santoro, E. Yacoub, and E. Formisano, "Representation of pitch chroma by multi-peak spectral tuning in human auditory cortex," *NeuroImage*, vol. 106, pp. 161–169, feb 2015.

[62] X. Wang, "The harmonic organization of auditory cortex.," *Frontiers in Systems Neuroscience*, vol. 7, no. December, p. 114, 2013.

[63] I. Bitterman, Yael and Mukamel, Roy and Malach, Rafael and Fried, Itzhak and Nelken, "Ultra-fine frequency tuning revealed in single neurons of human auditory cortex," *Nature*, vol. 451, no. 7175, pp. 197–201, 2008.

[64] D. Bendor, M. S. Osmanski, and X. Wang, "Dual-pitch processing mechanisms in primate auditory cortex.," *The Journal of Neuroscience*, vol. 32, pp. 16149–61, nov 2012.

[65] S. Norman-Haignere, N. Kanwisher, and J. H. McDermott, "Cortical pitch regions in humans respond primarily to resolved harmonics and are located in specific tonotopic regions of anterior auditory cortex.," *The Journal of Neuroscience*, vol. 33, no. 50, pp. 19451–69, 2013.

[66] D. Bendor and X. Wang, "The neuronal representation of pitch in primate auditory cortex.," *Nature*, vol. 436, no. 7054, pp. 1161–1165, 2005.

[67] K. M. M. Walker, J. K. Bizley, A. J. King, and J. W. H. Schnupp, "Multiplexed and robust representations of sound features in auditory cortex.," *Journal of Neuroscience*, vol. 31, no. 41, pp. 14565–14576, 2011.

[68] J. K. Bizley, K. M. M. Walker, F. R. Nodal, A. J. King, and J. W. H. Schnupp, "Auditory cortex represents both pitch judgments and the corresponding acoustic cues," *Current Biology*, vol. 23, no. 7, pp. 620–625, 2013.

[69] C. Suied, T. R. Agus, S. J. Thorpe, N. Mesgarani, and D. Pressnitzer, "Auditory gist: recognition of very short sounds from timbre cues.," *The Journal of the Acoustical Society of America*, vol. 135, no. 3, pp. 1380–91, 2014.

[70] I. Nelken, J. Bizley, S. A. Shamma, and X. Wang, "Auditory Cortical Processing in Real-World Listening: The Auditory System Going Real," *The Journal of Neuroscience*, vol. 34, no. 46, pp. 15135–15138, 2014.

[71] R. Ritsma and A. Hoekstra, "Frequency Selectivity and the Tonal Residue," in *Facts and Models in Hearing* (E. Zwicker and E. Terhardt, eds.), ch. 21, pp. 156–163, Springer Berlin Heidelberg, 1974.

[72] B. C. J. Moore, B. R. Glasberg, and M. J. Shailer, "Frequency and intensity diffference limens for harmonics within complex tones," *The Journal of the Acoustical Society of America*, vol. 75, no. 2, pp. 550–561, 1984.

[73] J. F. Schouten, "The perception of subjective tones," in *Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen*, pp. 1086—-1093, 1938.

[74] C. Kaernbach and L. Demany, "Psychophysical evidence against the autocorrelation theory," *Journal of the Acoustical Society of America*, vol. 104, no. 4, pp. 2298–2306, 1998.

[75] J. H. Grose, J. W. Hall, and E. Buss, "Virtual pitch integration for asynchronous harmonics," *The Journal of the Acoustical Society of America*, vol. 112, no. 6, p. 2956, 2002.

[76] S. Denham, "Pitch detection of dynamic iterated rippled noise by humans and a modified auditory model," *BioSystems*, vol. 79, no. 1-3 SPEC. ISS., pp. 199–206, 2005.

[77] D. Pressnitzer, R. D. Patterson, and K. Krumbholz, "The lower limit of melodic pitch.," *The Journal of the Acoustical Society of America*, vol. 109, no. 5 Pt. 1, pp. 2074–2084, 2001.

[78] A. H. Takeuchi and S. H. Hulse, "Absolute Pitch," *Psychological Bulletin*, vol. 113, no. 2, pp. 345–361, 1993.

[79] S. Murakami and Y. Okada, "Contributions of principal neocortical neurons to magnetoencephalography and electroencephalography signals," *The Journal of Physiology*, vol. 575, no. 3, pp. 925–936, 2006.

[80] S. Williamson and L. Kaufman, "Theory of neuroelectric and neuromagnetic fields," *Advances in audiology*, vol. 6, pp. 1–39, 1990.

[81] R. P. Feynman, R. B. Leighton, and M. Sands, *The Feynman Lectures on Physics, Vol. II.* Addison–Wesley, 2011.

[82] M. Scherg, "Fundamentals of dipole source potential analysis," *Advances in Audiology*, vol. 6, pp. 40–69, 1990.

[83] B. Lütkenhöner and D. Poeppel, "From Tones to Speech: Magnetoencephalographic Studies," in *The Auditory Cortex* (J. A. Winer and C. E. Schreiner, eds.), pp. 597–615, Boston, MA: Springer US, 2011.

[84] D. L. Schomer and F. L. da Silva, *Niedermeyer's Electroencephalography: Basic Principles, Clinical Applications, and Related Fields.* Lippincott Williams and Wilkins, 6th editio ed., 2012.

[85] A. J. Shahin, L. E. Roberts, L. M. Miller, K. L. McDonald, and C. Alain, "Sensitivity of EEG and MEG to the N1 and P2 auditory evoked responses modulated by spectral complexity of sounds.," *Brain Topography*, vol. 20, pp. 55–61, jan 2007.

[86] E. Skoe and N. Kraus, "Auditory brainstem reponse to complex sounds: a tutorial," *Ear Hear*, vol. 31, no. 3, pp. 302–324, 2010.

[87] A. Krishnan, "Human frequency following response," in *Auditory evoked potentials: Basic principles and clinical application* (R. F. Burkard, J. J. Eggermont, and M. Don, eds.), pp. 313–335, Lippincott Williams & Wilkins Baltimore, 2007.

[88] E. B. J. Coffey, S. C. Herholz, A. M. P. Chepesiuk, S. Baillet, and R. J. Zatorre, "Cortical contributions to the auditory frequency-following response revealed by MEG," *Nature Communications*, vol. 7, p. 11070, mar 2016.

[89] A. Rupp, S. Uppenkamp, A. Gutschalk, R. Beucker, R. D. Patterson, T. Dau, and M. Scherg, "The representation of peripheral neural activity in the middle-latency evoked field of primary auditory cortex in humans," *Hearing Research*, vol. 174, no. 1-2, pp. 19–31, 2002.

[90] M. Scherg, R. Hari, and M. Hämäläinen, "Frequency-specific sources of the auditory N19-P30-P50 response detected by a multiple source analysis of evoked magnetic fields and potentials," in *Advances in Biomagnetism*, pp. 97–100, Springer US, 1989.

[91] T. P. Roberts, P. Ferrari, S. M. Stufflebeam, and D. Poeppel, "Latency of the auditory evoked neuromagnetic field components: stimulus dependence and insights toward perception," *Journal of Clinical Neurophysiology*, vol. 17, no. 2, pp. 114–29, 2000.

[92] R. Näätänen and T. Picton, "The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure.," *Psychophysiology*, vol. 24, no. 4, pp. 375–425, 1987.

[93] C. Alain, D. L. Woods, and D. Covarrubias, "Activation of duration-sensitive auditory cortical fields in humans," *Electroencephalography and Clinical Neurophysiology - Evoked Potentials*, vol. 104, no. 6, pp. 531–539, 1997.

[94] M. Andermann, R. van Dinther, R. D. Patterson, and A. Rupp, "Neuromagnetic representation of musical register information in human auditory cortex.," *NeuroImage*, vol. 57, pp. 1499–506, aug 2011.

[95] H. Beagley and J. Knight, "Changes in auditory evoked response with intensity," *The Journal of Laryngology & Laryngology Otology*, vol. 81, no. 8, pp. 861–873, 1967.

[96] A. Rupp and S. Uppenkamp, "Neuromagnetic Representation of Short Melodies in the Auditory Cortex," in *NAG/DAGA International Conference*, pp. 473–474, 2005.

[97] T. Rosburg, K. Zimmerer, and R. Huonker, "Short-term habituation of auditory evoked potential and neuromagnetic field components in dependence of the interstimulus interval.," *Experimental Brain Research*, vol. 205, pp. 559–70, oct 2010.

[98] H. Okamoto, H. Stracke, P. Bermudez, and C. Pantev, "Sound Processing Hierarchy within Human Auditory Cortex," *Journal of Cognitive Neuroscience*, vol. 23, no. 8, pp. 1855–1863, 2011.

[99] A. Seither-Preisler, R. Patterson, K. Krumbholz, S. Seither, and B. Lütkenhöner, "Evidence of pitch processing in the N100m component of the auditory evoked field.," *Hearing Research*, vol. 213, pp. 88–98, mar 2006.

[100] K. E. Crowley and I. M. Colrain, "A review of the evidence for P2 being an independent component process: age, sleep and modality.," *Clinical Neurophysiology*, vol. 115, pp. 732–44, apr 2004.

[101] R. D. Patterson, S. Uppenkamp, M. Andermann, and A. Rupp, "Brain imaging the activity associated with pitch intervals in a melody," in *167th Meeting of the Acoustical Society of America*, p. 2348, Journal of the Acoustical Society of America, 2014.

[102] I. Choi, H. M. Bharadwaj, S. Bressler, P. Loui, K. Lee, and B. G. Shinn-Cunningham, "Automatic processing of abstract musical tonality.," *Frontiers in Human Neuroscience*, vol. 8, no. December, p. 988, 2014.

[103] B. Lütkenhöner, A. Seither-Preisler, and S. Seither, “Piano tones evoke stronger magnetic fields than pure tones or noise, both in musicians and non-musicians.,” *NeuroImage*, vol. 30, pp. 927–37, apr 2006.

[104] A. Shahin, L. E. Roberts, C. Pantev, L. J. Trainor, and B. Ross, “Modulation of P2 auditory-evoked responses by the spectral complexity of musical sounds.,” *Neuroreport*, vol. 16, pp. 1781–5, dec 2005.

[105] A. Gutschalk, R. D. Patterson, A. Rupp, S. Uppenkamp, and M. Scherg, “Sustained magnetic fields reveal separate sites for sound level and temporal regularity in human auditory cortex.,” *NeuroImage*, vol. 15, pp. 207–16, jan 2002.

[106] A. Gutschalk, R. D. Patterson, M. Scherg, S. Uppenkamp, and A. Rupp, “Temporal dynamics of pitch in human auditory cortex,” *NeuroImage*, vol. 22, no. 2, pp. 755–766, 2004.

[107] A. Gutschalk, R. D. Patterson, S. Uppenkamp, M. Scherg, and A. Rupp, “Recovery and refractoriness of auditory evoked fields after gaps in click trains.,” *The European Journal of Neuroscience*, vol. 20, pp. 3141–7, dec 2004.

[108] A. Gutschalk, R. D. Patterson, M. Scherg, S. Uppenkamp, and A. Rupp, “The effect of temporal context on the sustained pitch response in human auditory cortex.,” *Cerebral Cortex*, vol. 17, pp. 552–61, mar 2007.

[109] P. Schneider, M. Scherg, H. G. Dosch, H. J. Specht, A. Gutschalk, and A. Rupp, “Morphology of Heschl’s gyrus reflects enhanced activation in the auditory cortex of musicians.,” *Nature Neuroscience*, vol. 5, pp. 688–94, jul 2002.

[110] G. Musacchia, D. Strait, and N. Kraus, “Relationships between behavior, brainstem and cortical encoding of seen and heard speech in musicians and non-musicians,” *Hearing Research*, vol. 241, no. 1-2, pp. 34–42, 2008.

[111] K. Tremblay, N. Kraus, T. McGee, C. Ponton, and B. Otis, “Central auditory plasticity: changes in the N1-P2 complex after speech-sound training.,” *Ear and Hearing*, vol. 22, pp. 79–90, may 2001.

[112] E. M. O. Borchert, C. Micheyl, and A. J. Oxenham, “Perceptual Grouping Affects Pitch Judgments Across Time and Frequency,” *Journal of Experimental Psychology: Human Perception and Performance*, vol. 37, no. 1, pp. 257–269, 2011.

[113] M. S. A. Zilany and I. C. Bruce, “Modeling auditory-nerve responses for high sound pressure levels in the normal and impaired auditory periphery,” *The Journal of the Acoustical Society of America*, vol. 120, no. 3, p. 1446, 2006.

[114] T. Irino and R. Patterson, “A time-domain, level-dependent auditory filter: The gammachirp,” *The Journal of the Acoustical Society of America*, vol. 101, no. 1, pp. 412–419, 1997.

[115] T. Irino and R. D. Patterson, “A compressive gammachirp auditory filter for both physiological and psychophysical data,” *The Journal of the Acoustical Society of America*, vol. 109, pp. 2008–2022, may 2001.

[116] E. Lopez-Poveda and R. Meddis, “A human nonlinear cochlear filterbank,” *The Journal of the Acoustical Society of America*, vol. 110, no. 6, pp. 3170–3118, 2001.

[117] M. S. A. Zilany, I. C. Bruce, P. C. Nelson, and L. H. Carney, "A phenomenological model of the synapse between the inner hair cell and auditory nerve: long-term adaptation with power-law dynamics.," *The Journal of the Acoustical Society of America*, vol. 126, no. 5, pp. 2390–412, 2009.

[118] C. Giguere and P. Woodland, "A computational model of the auditory periphery for speech and hearing research. II. Descending paths," *The Journal of the Acoustical Society of America*, vol. 95, no. January, pp. 343–349, 1994.

[119] A. de Cheveigné, "Pitch Perception Models," in *Pitch: Neural Coding and Perception* (C. J. Plack, R. R. Fay, A. J. Oxenham, and A. N. Popper, eds.), ch. 6, pp. 169–233, Springer New York, 2005.

[120] L. Gao, K. Kostlan, Y. Wang, and X. Wang, "Distinct Subthreshold Mechanisms Underlying Rate-Coding Principles in Primate Auditory Cortex," *Neuron*, vol. 91, pp. 905–919, aug 2016.

[121] S. Shamma and D. Klein, "The case of the missing pitch templates: How harmonic templates emerge in the early auditory system," *The Journal of the Acoustical Society of America*, vol. 107, no. 5, pp. 2631–2644, 2000.

[122] M. A. Cohen, S. Grossberg, and L. L. Wyse, "A spectral network model of pitch perception.," *Journal of the Acoustical Society of America*, vol. 98, no. 2 Pt 1, pp. 862–879, 1995.

[123] J. C. R. Licklider, "A duplex theory of pitch perception.," *Experientia*, vol. 7, no. 4, pp. 128–34, 1951.

[124] R. Meddis and L. O'Mard, "A unitary model of pitch perception.," *The Journal of the Acoustical Society of America*, vol. 102, pp. 1811–1820, sep 1997.

[125] E. Balaguer-Ballester, S. L. Denham, and R. Meddis, "A cascade autocorrelation model of pitch perception.," *The Journal of the Acoustical Society of America*, vol. 124, pp. 2186–95, oct 2008.

[126] E. Balaguer-Ballester, M. Coath, and S. L. Denham, "A model of perceptual segregation based on clustering the time series of the simulated auditory nerve firing probability," *Biological Cybernetics*, vol. 97, pp. 479–491, dec 2007.

[127] L. Wiegrebe and R. Meddis, "The Representation of Periodic Sounds in Simulated Sustained Chopper Units of the Ventral Cochlear Nucleus," *The Journal of the Acoustical Society of America*, vol. 115, no. 3, pp. 1207–1218, 2004.

[128] R. D. Patterson, "The sound of a sinusoid: Spectral models," *The Journal of the Acoustical Society of America*, vol. 96, no. 3, p. 1409, 1994.

[129] N. Erfanian Saeedi, P. J. Blamey, A. N. Burkitt, and D. B. Grayden, "Learning Pitch with STDP: A Computational Model of Place and Temporal Pitch Perception Using Spiking Neural Networks," *PLoS Computational Biology*, vol. 12, p. e1004860, apr 2016.

[130] W. Gerstner, R. Kempter, J. L. van Hemmen, and H. Wagner, "A neuronal learning rule for sub-millisecond temporal coding," *Nature*, vol. 383, pp. 76–78, sep 1996.

[131] L. Wiegrebe, "Searching for the time constant of neural pitch extraction," *The Journal of the Acoustical Society of America*, vol. 109, pp. 1082–1091, mar 2001.

[132] R. D. Patterson, M. H. Allerhand, and C. Giguère, “Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform,” *The Journal of the Acoustical Society of America*, vol. 98, p. 1890, oct 1995.

[133] S. J. Kiebel, M. I. Garrido, R. J. Moran, and K. J. Friston, “Dynamic causal modelling for EEG and MEG.,” *Cognitive neurodynamics*, vol. 2, pp. 121–36, jun 2008.

[134] C. C. Kerr, C. J. Rennie, and P. A. Robinson, “Physiology-based modeling of cortical auditory evoked potentials.,” *Biological Cybernetics*, vol. 98, no. 2, pp. 171–84, 2008.

[135] R. D. Patterson, “The sound of a sinusoid: Time-interval models,” *The Journal of the Acoustical Society of America*, vol. 96, no. 3, p. 1419, 1994.

[136] W. A. Yost, R. Patterson, and S. Sheft, “A time domain description for the pitch strength of iterated rippled noise,” *The Journal of the Acoustical Society of America*, vol. 99, no. 2, p. 1066, 1996.

[137] M. Andermann, R. D. Patterson, M. Geldhauser, N. Sieroka, and A. Rupp, “Duifhuis Pitch: Neuromagnetic representation and auditory modeling.,” *Journal of Neurophysiology*, aug 2014.

[138] K. J. Friston, L. M. Harrison, and W. Penny, “Dynamic causal modelling.,” *NeuroImage*, vol. 19, pp. 1273–302, aug 2003.

[139] K. Friston, “Causal Modelling and Brain Connectivity in Functional Magnetic Resonance Imaging,” *PLoS Biology*, vol. 7, p. e1000033, feb 2009.

[140] S. D. Kiebel, R. J. Moran, H. E. den Ouden, J. Daunizeau, and K. J. Friston, “Ten simple rules for dynamic causal modeling,” *NeuroImage*, vol. 49, pp. 3099–3109, feb 2010.

[141] J. Daunizeau, O. David, and K. E. Stephan, “Dynamic causal modelling: A critical review of the biophysical and statistical foundations,” *NeuroImage*, vol. 58, pp. 312–322, sep 2011.

[142] J. Daunizeau, S. J. Kiebel, and K. J. Friston, “Dynamic causal modelling of distributed electromagnetic responses,” *NeuroImage*, vol. 47, pp. 590–601, aug 2009.

[143] R. Moran, D. a. Pinotsis, and K. Friston, “Neural masses and fields in dynamic causal modeling,” *Frontiers in Computational Neuroscience*, vol. 7, p. 57, jan 2013.

[144] K. Friston, J. Mattout, N. Trujillo-Barreto, J. Ashburner, and W. Penny, “Variational free energy and the Laplace approximation.,” *NeuroImage*, vol. 34, pp. 220–34, jan 2007.

[145] A. C. Marreiros, S. J. Kiebel, J. Daunizeau, L. M. Harrison, and K. J. Friston, “Population dynamics under the Laplace assumption.,” *NeuroImage*, vol. 44, pp. 701–14, feb 2009.

[146] H. A. David, *The method of paired comparisons.* New York: Oxford University Press, 1988.

[147] R. D. Patterson and T. Irino, “Modeling temporal asymmetry in the auditory system.,” *The Journal of the Acoustical Society of America*, vol. 104, pp. 2967–79, nov 1998.

[148] M. Carandini and D. J. Heeger, “Normalization as a canonical neural computation.,” *Nature Reviews. Neuroscience*, vol. 13, no. 1, pp. 51–62, 2011.

[149] P. M. Briley, C. Breakey, and K. Krumbholz, "Evidence for Pitch Chroma Mapping in Human Auditory Cortex," *Cerebral Cortex*, vol. 23, pp. 2601–2610, nov 2013.

[150] K. Wimmer, A. Compte, A. Roxin, D. Peixoto, A. Renart, and J. de la Rocha, "Sensory integration dynamics in a hierarchical network explains choice probabilities in cortical area MT.," *Nature Communications*, vol. 6, p. 6177, 2015.

[151] K.-F. Wong and X.-J. Wang, "A recurrent network mechanism of time integration in perceptual decisions.," *The Journal of Neuroscience*, vol. 26, no. 4, pp. 1314–1328, 2006.

[152] S. Zschocke and H.-C. Hansen, "Entstehungsmechanismen des EEG," in *Klinische Elektroenzephalographie* (S. Zschocke and H.-C. Hansen, eds.), pp. 1–29, Berlin, Heidelberg: Springer Berlin Heidelberg, 2012.

[153] L. Feng and X. Wang, "Harmonic template neurons in primate auditory cortex underlying complex sound processing," *Proceedings of the National Academy of Sciences*, p. 201607519, 2017.

[154] W. Gerstner and W. M. Kistler, *Spiking Neuron Models*, vol. 66. Cambridge University Press, 2002.

[155] X. J. Wang, "Probabilistic decision making by slow reverberation in cortical circuits," *Neuron*, vol. 36, no. 5, pp. 955–968, 2002.

[156] S. Ostojic and N. Brunel, "From spiking neuron models to linear-nonlinear models.," *PLoS computational biology*, vol. 7, no. 1, p. e1001056, 2011.

[157] N. Fourcaud-Trocmé, D. Hansel, C. van Vreeswijk, and N. Brunel, "How spike generation mechanisms determine the neuronal response to fluctuating inputs," *The Journal of Neuroscience*, vol. 23, no. 37, pp. 11628–11640, 2003.

[158] M. J. E. Richardson, "Firing-rate response of linear and nonlinear integrate-and-fire neurons to modulated current-based and conductance-based synaptic drive," *Physical Review E*, vol. 76, no. 2, p. 021919, 2007.

[159] K. D. Miller and F. Fumarola, "Mathematical Equivalence of Two Common Forms of Firing Rate Models of Neural Networks," *Neural Computation*, vol. 24, no. 1, pp. 25–31, 2012.

[160] N. Brunel and X. J. Wang, "Effects of neuromodulation in a cortical network model of object working memory dominated by recurrent inhibition," *Journal of Computational Neuroscience*, vol. 11, no. 1, pp. 63–85, 2001.

[161] G. Deco, A. Ponce-Alvarez, D. Mantini, G. L. Romani, P. Hagmann, and M. Corbetta, "Resting-State Functional Connectivity Emerges from Structurally and Dynamically Shaped Slow Linear Fluctuations," *The Journal of Neuroscience*, vol. 33, pp. 11239–11252, jul 2013.

[162] K. Friston, "A theory of cortical responses.," *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, vol. 360, pp. 815–36, apr 2005.

[163] M. S. A. Zilany and L. H. Carney, "Power-Law Dynamics in an Auditory-Nerve Model Can Account for Neural Adaptation to Sound-Level Statistics," *Journal of Neuroscience*, vol. 30, pp. 10380–10390, aug 2010.

[164] R. D. Patterson and F. L. Wightman, “Residue pitch as a function of component spacing.,” *The Journal of the Acoustical Society of America*, vol. 59, no. 6, pp. 1450–1459, 1976.

[165] D. Bendor, “The Role of Inhibition in a Computational Model of an Auditory Cortical Neuron during the Encoding of Temporal Information,” *PLoS computational biology*, vol. 11, no. 4, p. e1004197, 2015.

[166] P. Friedel and J. L. Van Hemmen, “Inhibition, not excitation, is the key to multimodal sensory integration,” *Biological Cybernetics*, vol. 98, no. 6, pp. 597–618, 2008.

[167] S. Denève and C. K. Machens, “Efficient codes and balanced networks,” *Nature Neuroscience*, vol. 19, pp. 375–382, feb 2016.

[168] K. Friston and S. Kiebel, “Predictive coding under the free-energy principle.,” *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, vol. 364, no. 1521, pp. 1211–21, 2009.

[169] K. Rauss and G. Pourtois, “What is bottom-up and what is top-down in predictive coding,” *Frontiers in Psychology*, vol. 4, pp. 1–8, 2013.

[170] R. Parncutt and G. Hair, “Consonance and dissonance in music theory and psychology: Disentangling dissonant dichotomies,” *Journal of Interdisciplinary Music Studies*, vol. 5, no. 2, pp. 119–166, 2011.

[171] I. Peretz, A. J. Blood, V. Penhune, and R. Zatorre, “Cortical deafness to dissonance,” *Brain*, vol. 124, pp. 928–940, may 2001.

[172] R. Plomp and W. J. M. Levelt, “Tonal Consonance and Critical Bandwidth,” *The Journal of the Acoustical Society of America*, vol. 38, no. 4, pp. 548–560, 1965.

[173] K. Itoh, S. Suwazono, and T. Nakada, “Central auditory processing of noncontextual consonance in music: an evoked potential study.,” *The Journal of the Acoustical Society of America*, vol. 128, no. 6, pp. 3781–3787, 2010.

[174] O. Bones and C. J. Plack, “Subcortical representation of musical dyads: Individual differences and neural generators,” *Hearing Research*, vol. 323, pp. 9–21, 2015.

[175] G. M. Bidelman and J. Grall, “Functional organization for musical consonance and tonal pitch hierarchy in human auditory cortex,” *NeuroImage*, vol. 101, pp. 204–214, 2014.

[176] G. M. Bidelman, “The Role of the Auditory Brainstem in Processing Musically Relevant Pitch,” *Frontiers in Psychology*, vol. 4, no. MAY, pp. 1–13, 2013.

[177] P. Regnault, E. Bigand, and M. Besson, “Different brain mechanisms mediate sensitivity to sensory consonance and harmonic context: evidence from auditory event-related brain potentials.,” *Journal of Cognitive Neuroscience*, vol. 13, no. 2, pp. 241–255, 2001.

[178] M. Kolinski, “Consonance and Dissonance,” *Ethnomusicology*, vol. 6, p. 66, may 1962.

[179] A. Kameoka and M. Kuriyagawa, “Consonance theory part I: consonance of dyads.,” *The Journal of the Acoustical Society of America*, vol. 45, no. February, pp. 1451–1459, 1969.

[180] D. A. Schwartz, C. Q. Howe, and D. Purves, “The statistical structure of human speech sounds predicts musical universals.,” *The Journal of Neuroscience*, vol. 23, no. 18, pp. 7160–7168, 2003.

[181] G. M. Bidelman and A. Krishnan, "Neural Correlates of Consonance, Dissonance, and the Hierarchy of Musical Pitch in the Human Brainstem," *Journal of Neuroscience*, vol. 29, pp. 13165–13171, oct 2009.

[182] G. M. Bidelman and A. Krishnan, "Brainstem correlates of behavioral and compositional preferences of musical harmony," *NeuroReport*, vol. 22, pp. 212–216, mar 2011.

[183] L. Minati, C. Rosazza, L. D'Incerti, E. Pietrocini, L. Valentini, V. Scaioli, C. Loveday, and M. G. Bruzzone, "Functional MRI/event-related potential study of sensory consonance and dissonance in musicians and nonmusicians.," *Neuroreport*, vol. 20, no. 1, pp. 87–92, 2009.

[184] D. Schön, P. Regnault, S. Ystad, and M. Besson, "Sensory Consonance," *Music Perception*, vol. 23, pp. 105–118, dec 2005.

[185] M. J. Tramo, P. A. Cariani, B. Delgutte, and L. D. Braida, "Neurobiological Foundations for the Theory of Harmony in Western Tonal Music," *Annals of the New York Academy of Sciences*, vol. 930, pp. 92–116, jan 2006.

[186] J. H. Mcdermott, A. F. Schultz, E. A. Undurraga, and R. A. Godoy, "Indifference to dissonance in native Amazonians reveals cultural variation in music perception," *Nature*, vol. 535, no. 7613, pp. 547–550, 2016.

[187] J. W. Butler and P. G. Daston, "Musical consonance as musical preference: A cross-cultural study.," *The Journal of General Psychology*, vol. 79, no. February 2015, pp. 129–142, 1968.

[188] L. J. Trainor, C. D. Tsang, and V. H. W. Cheung, "Preference for sensory consonance in 2- and 4-month-old infants," *Music Perception*, vol. 20, pp. 187–194, dec 2002.

[189] J. Plantinga and S. E. Trehub, "Revisiting the Innate Preference for Consonance.," *Journal of Experimental Psychology. Human Perception and Performance*, vol. 40, no. 1, pp. 40–49, 2014.

[190] J. H. McDermott, A. J. Lehr, and A. J. Oxenham, "Individual differences reveal the basis of consonance," *Current Biology*, vol. 20, no. 11, pp. 1035–1041, 2010.

[191] F. Krueger, "Consonance and Dissonance," *The Journal of Philosophy, Psychology and Scientific Methods*, vol. 10, p. 158, mar 1913.

[192] Y. I. Fishman, I. O. Volkov, M. D. Noh, P. C. Garell, H. Bakken, J. C. Arezzo, M. A. Howard, and M. Steinschneider, "Consonance and dissonance of musical chords: neural correlates in auditory cortex of monkeys and humans.," *Journal of Neurophysiology*, vol. 86, pp. 2761–88, dec 2001.

[193] W. A. Sethares, "Local consonance and the relationship between timbre and scale," *The Journal of the Acoustical Society of America*, vol. 94, p. 1218, nov 1993.

[194] A. Kameoka and M. Kuriyagawa, "Consonance Theory Part II: Consonance of Complex Tones and Its Calculation Method," *The Journal of the Acoustical Society of America*, vol. 45, no. 6, p. 1460, 1969.

[195] M. Cousineau, J. H. McDermott, and I. Peretz, "The basis of musical consonance as revealed by congenital amusia.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, no. 48, pp. 19858–63, 2012.

[196] M. Ebeling, “Neuronal periodicity detection as a basis for the perception of consonance: a mathematical model of tonal fusion,” *Journal of the Acoustical Society of America*, vol. 124, pp. 2320–2329, oct 2008.

[197] T. H. Fritz, W. Renders, K. Müller, P. Schmude, M. Leman, R. Turner, and A. Villringer, “Anatomical differences in the human inferior colliculus relate to the perceived valence of musical consonance and dissonance,” *European Journal of Neuroscience*, pp. n/a–n/a, jul 2013.

[198] B. Efron and R. LePage, “Introduction to bootstrap,” *Exploring the limits of bootstrap*, pp. 3–10, 1992.

[199] E. Gordon, *Introduction to research and the psychology of music.* GIA, 1998.

[200] S. Kumar and M. Schönwiesner, “Mapping human pitch representation in a distributed system using depth-electrode recordings and modeling.,” *The Journal of Neuroscience*, vol. 32, pp. 13348–51, sep 2012.

[201] C. C. Wier, W. Jesteadt, and D. M. Green, “Intensity discrimination as a function of frequency and sensation level,” *The Journal of the Acoustical Society of America*, vol. 61, no. 1, pp. 169–177, 1977.

[202] G. Tononi, “An information integration theory of consciousness,” *BMC Neuroscience*, vol. 5, no. 1, 2004.

[203] C. Huang, B. Englitz, S. Shamma, and J. Rinzel, “A neuronal network model for context-dependence of pitch change perception,” *Frontiers in Computational Neuroscience*, vol. 9, p. 101, aug 2015.

[204] P. Loui, H. C. Li, A. Hohmann, and G. Schlaug, “Enhanced Cortical Connectivity in Absolute Pitch Musicians: A Model for Local Hyperconnectivity,” *Journal of Cognitive Neuroscience*, vol. 4, no. 23, pp. 1015–1026, 2010.

[205] M. Wengenroth, M. Blatow, A. Heinecke, J. Reinhardt, C. Stippich, E. Hofmann, and P. Schneider, “Increased Volume and Function of Right Auditory Cortex as a Marker for Absolute Pitch,” *Cerebral Cortex*, vol. 24, pp. 1127–1137, 2014.

[206] J. B. Fritz, M. Elhilali, S. V. David, and S. A. Shamma, “Auditory attention—focusing the searchlight on sound,” *Current Opinion in Neurobiology*, vol. 17, pp. 437–455, aug 2007.

[207] B. Scharf, S. Quigley, C. Aoki, N. Peachey, and A. Reeves, “Focused auditory attention and frequency selectivity,” *Perception & Psychophysics*, vol. 42, pp. 215–223, may 1987.

[208] T. Folyi, B. Fehér, and J. Horváth, “Stimulus-focused attention speeds up auditory processing,” *International Journal of Psychophysiology*, vol. 84, no. 2, pp. 155–163, 2012.

[209] F. Baluch and L. Itti, “Mechanisms of top-down attention,” *Trends in Neurosciences*, vol. 34, no. 4, pp. 210–224, 2011.

[210] S. Shamma and J. Fritz, “Adaptive auditory computations,” *Current Opinion in Neurobiology*, vol. 25, pp. 164–168, 2014.

[211] E. Bigand and B. Tillmann, “Effect of Context on the Perception of Pitch Structures,” in *Pitch*, pp. 306–351, New York: Springer-Verlag, 2000.

Appendix A

Tuning procedure for the parameters of the cortical model

After fixing the structure of the connectivity weight matrices and the normalisation parameters of the subcortical input, the dynamics of the cortical model still depend on 35 different parameters. Parameters values were chosen following a five-stages procedure. This appendix details the criteria used to fix the value of those parameters at each of the states (see also Table 4.1).

Fixed parameters 17 cortical parameters were fixed ad-hoc using the values from the literature, without any further tuning.

Parameters of the transfer functions $\phi(I)$, a, b and d (see Equation 4.4) for excitatory and inhibitory ensembles, and the AMPA conductivities J^{**}_{AMPA} and $\hat{J}^{**}_{\text{AMPA}}$, were all taken from the original publication by Wong and Wang describing the ensemble rate model [151]

GABA and AMPA synaptic time constants, τ_{GABA} and τ_{AMPA}, and the NMDA coupling constant γ, were all taken from the original publication by Brunel and Wang describing the synaptic gating models [160]. The NMDA decay was set ad-hoc to $\tau_{\text{NDMA}} = 20\,\text{ms}$, within the typical range of this constant [162].

The adaptation time constant was set to $\tau_{\text{adap}} = 100\,\text{ms}$, according to the literature [154].

Stage 1: decoder's sensitivity to input The rest of the constants were initialised to zero and progressively tuned using subcortical input generated with iterated rippled noises along five consecutive stages. At each stage, we focused on a particular set of parameters, keeping the values of the already tuned parameters fixed and the values of the parameters to be tuned during consequent stages equal to zero.

First, we used the normalised SACF of two different IRNs with 16 iterations and delays $d = 8\,\text{ms}$ and $d = 5\,\text{ms}$ to adjust the sensitivity of the decoder to subcortical input by tuning the thalamic conductivity J^{th}_{AMPA}, the NMDA self-excitation J^{ee}_{NDMA}, the ground population time constant τ^{0}_{pop}, and the baseline excitatory input I^{e}_{0}.

Parameters were first set to baseline values from the literature [151] and then tuned to make the decoder excitatory ensembles sensitive enough as to capture the peaks of the SACF input, but robust enough as to ignore spurious noisy activity in the subcortical input.

Then, we tuned the adaptation strength α so that populations could not reach firing rates above H_e ~200-300 Hz.

Stage 2: inhibitory build up at the decoder Next, we adjusted the inhibitory build up in response to the excitatory activity in the decoder, using the same stimuli. In this stage we fixed the conductivities towards inhibitory ensembles J^{ei}_{NDMA} and J^{ii}_{GABA}, and the inhibitory baseline input I^i_0. As in stage 1, we first initialised the values according to the literature [151], and then tuned the values to maximise the speed of the inhibitory build up at the populations encoding the pitch value, whilst minimising spurious activation at other inhibitory ensembles.

Stage 3: representation build up at the sustainer Then, we set the sustainer's parameters to allow the propagation of the pitch representation towards the higher-level network. First, we adjusted the baseline in put I^{sus}_0 and the connectivities at the sustainer $\hat{J}^{ee}_{\text{NMDA}}$, $\hat{J}^{ei}_{\text{NMDA}}$, $\hat{J}^{ie}_{\text{GABA}}$ and $\hat{J}^{ii}_{\text{GABA}}$, to ensure that the inhibitory ensembles were dominant, effectively shutting down any activation in the excitatory populations in absence of afferent input. For simplicity, conductivities targeting inhibitory ensembles, $\hat{J}^{ei}$ and $\hat{J}^{ii}$, were set to zero.

Next, we fixed the afferent conductivities $\hat{J}^a_{\text{AMPA}}$ and $\hat{J}^a_{\text{GABA}}$, and fine tuned the previous sustainer's conductivities, in order to guarantee that a joint excitatory and inhibitory activation at a given column at the decoder elicits a robust activation at the analogous excitatory ensemble at the sustainer.

Stage 4: decoding and sustainer replacement The next step was to adjust the inhibitory-to-excitatory conductivity J^{ie}_{GABA} at the decoder to the minimum value guaranteeing the inhibition of the peaks at the lower harmonics in the excitatory ensembles. The baseline inhibition weight c^{ie}_0 was then adjusted using the dyads to ensure that non-harmonically related peaks could coexist in the excitatory representation at the decoder.

Afterwards, we set the efferent NMDA connectivity J^e_{NMDA} so that the top-down input to the inhibitory ensembles in the decoder would replace the excitatory input coming from the lower harmonics after their inhibition.

Fine tuning Non ad-hoc parameters were further fine tuned using a wider array of stimuli, comprising 8 different IRNs with delays ranging from $d = 1\,\text{ms}$ to $d = 8\,\text{ms}$, and 6 different IRN dyads (minor second, third, fourth, tritone, fifth, and seventh) with a ground delay $d = 8\,\text{ms}$ (see also §5.3.2.1). In this last stage we ensured that, after the decoding, both decoder and sustainer show one (two) prominent peak(s) of activation corresponding to the pitch value(s) evoked by each stimulus.

Most of the fitted parameters accepted moderate perturbations of a 10%–20% of their final values without substantially compromising the dynamics of the model.

Parameter validation Parameters were validated using freshly sampled IRNs with 8, 16 and 32 iterations, harmonic complex tones with and without missing fundamentals, and pure tones. A range of pitch values between 200 Hz and 1000 Hz was considered for all stimulus types. After a transient state of around $100 - 150\,\text{ms}$, the activity in the decoder systematically converged to a state of equilibrium consisting on a unimodal distribution centred on the population corresponding to the pitch typically elicited by each class of stimulus (see Figure 4.4), in line with predictions of abstract pitch perception models [21]. Perceptual results for the different stimuli are shown in §4.3.1.

Appendix B

Horizontal pitch interactions

Despite the large evidence of melodic content and interval recognition being processed by higher cognitive centres in the ascending auditory pathway [171–173, 211], it is worth testing the effect of horizontal interactions on processing time. In this section, we will see that there are interactions between the harmonic series of different pitch values that can interact horizontally, affecting processing time in a noticeable way.

Pitch onset and offset in tone intervals

First, we studied if the timings of pitch processing were affected by the presence of a previous pitch value in the cortical representation. First, we run a series of simulations considering IRN intervals consisting on a fundamental tone $f_0 = 160$ transitioning to different notes within the chromatic scale. IRN tone shifts were mediated by 5 ms Hanning windows to ensure a smooth transition without changes in the power spectrum of the stimulation.

Dynamics of the inhibitory ensembles at the decoder during the transition were compared with the dynamics during the onset of the second tone (see Figure B.1) and the offset of the first tone. The onset was systematically delayed between 5 ms and 10 ms when the tone was preceded by another tone with respect to the silence-to-tone condition (see Figure B.2B). Offset of the previous tone was hastened between 25 ms and 30 ms for all transitions except the octave when the tone was followed by a second tone with respect to the tone-to-silence condition (see Figure B.2A).

Induced delay onsets and offsets were contrasted with the perceived consonance/dissonance sensation elicited by dyads comprising the notes of the intervals. Weak correlations were found between the two quantities.

Pitch transitions revisited

In §**??** we described the transition dynamics of the model in response to pitch changes in the stimulus as a reconfiguration of the attractor state and the associated transition of the cortical dynamics to the new basis of attraction. A mechanistic analysis of such transition reveals that the transient dynamics do depend on the harmonic relationship between the fundamental frequencies of the transitioning tones, as evidenced in Figure B.2.

Consider an arbitrary transition induced by a pitch change $T \rightarrow T'$ in the stimulation. The system first lies in an equilibrium state characterised by a high activation in the populations n characterising the first pitch value $\delta t_n = T$. After the transition the system will rest in a new equilibrium state with a large activity in populations n' such that $\delta t_{n'} = T'$.

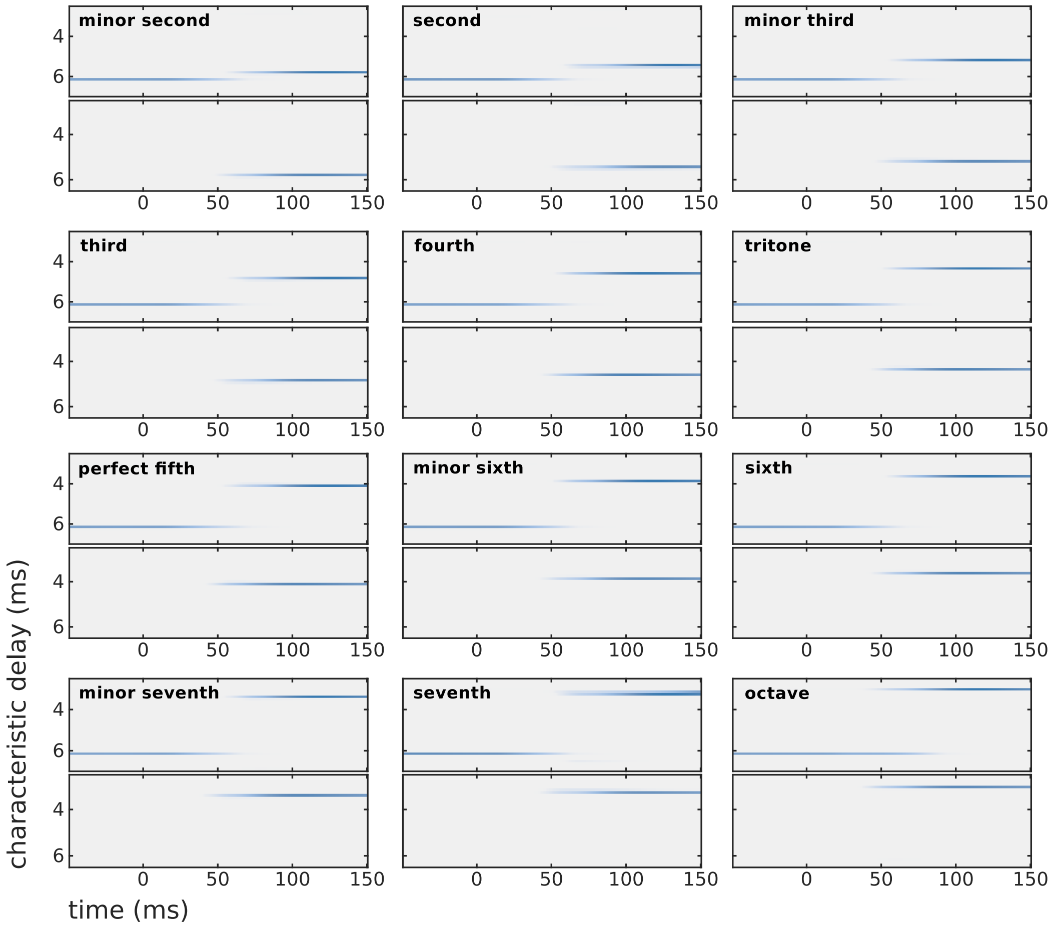


Figure B.1: Transition dynamics during pitch change for different chromatic intervals. Each panel compares the transition dynamics of the inhibitory ensembles at the decoder during pitch change with the onset (i.e. transition from silence) of the second tone. Note that the onset of the second note is, in general, delayed around 10 ms when preceded by another tone. The first tone was set to $f_0 = 160\,Hz$ and the second tone corresponds to each of the notes of the chromatic scale according to the tempered tuning (see §5.1.1). Stimulus onset/transition was set to $t = 0\,ms$. Stimuli were 16 iterations IRNs, bandpass filtered between 0.8 kHz and 3.2 kHz. Transitions were mediated by a 5 ms Hanning window. Simulations were performed without cortical noise in order to obtain a more robust measure of the effect.

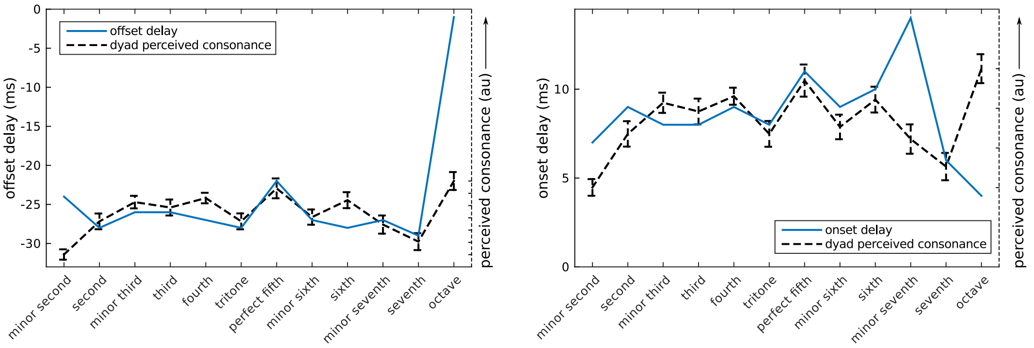


Figure B.2: Onset and offset delay induced by horizontal interactions. Plots quantitatively measure the delay induced by pitch interactions in the onset/offset of the neural representation in the inhibitory ensembles at the decoder. Perceptual measurements of the corresponding dyads are shown for reference. A) Offset delay: negative values indicate that the offset was faster when the tone was followed by a second tone than when the tone was followed by silence. B) Onset delay: positive values indicate that the onset was delayed when preceded by a previous tone than when preceded by silence. Measurements were taken from simulations shown in Figure B.1.

The transition dynamics between both states depend on the harmonic relationship between T and T'. If the fundamental periods are not harmonically related, the new subcortical input does not elicit activation in the decoder's excitatory ensemble encoding the previous pitch n, whose activity decreases rapidly. The inhibitory ensemble n, in the other hand, presents a larger inertia due to the reinforcement of the sustainer, and stays active for a longer period. However, the new elicited activity in the excitatory populations at the decoder will trigger a new decoding process, effectively activating the inhibitory ensemble n'. Due to the non-zero connections between inhibitory ensembles in the decoder (see Figure 4.2D), the activation of the inhibitory ensemble n' will accelerate the shunting of the inhibitory ensemble n, effectively accelerating the offset of the previous pitch representation (see Figure B.2B).

Moreover, since the inhibitory ensemble n is still active right after the pitch change has been reflected upon the subcortical input, global inhibition will slow down the build up of the new SACF representation in the decoders' excitatory populations, effectively delaying the onset of the decoding process. Due to the increased inhibition of n toward populations encoding lower harmonics, this effect is larger when T' shares any lower harmonic with T, as in the case of the perfect fifth (see Figure B.2B).

If the fundamental periods of the two tones in the transition are harmonically related, the dynamics change subtly. When $T < T'$, the excitatory input characterising the new pitch value T' reinforces the excitatory ensemble n at the decoder, which thus effectively reinforces the activity of the inhibitory ensemble n slowing down its shunting process (see the octave in Figure B.2A). Since the excitatory input at n is already active on the onset of the new tone, it effectively contributes to the build up of the inhibitory ensemble at n', and the decoding process is triggered earlier (see the octave in Figure B.2B).

When $T > T'$, the situation is reversed (see Figure B.3). The active inhibitory ensemble at n effectively inhibits the excitation at n', slowing down the excitatory build up of the new tone and thus increasing the onset delay. Due to the slow rise of the inhibitory activity at n', the offset speed is not substantially altered. In this last case, the transitions result in a less salient response, consistent with EEG recordings reporting an increased adaptation effect when transitioning between harmonically related IRNs [149].

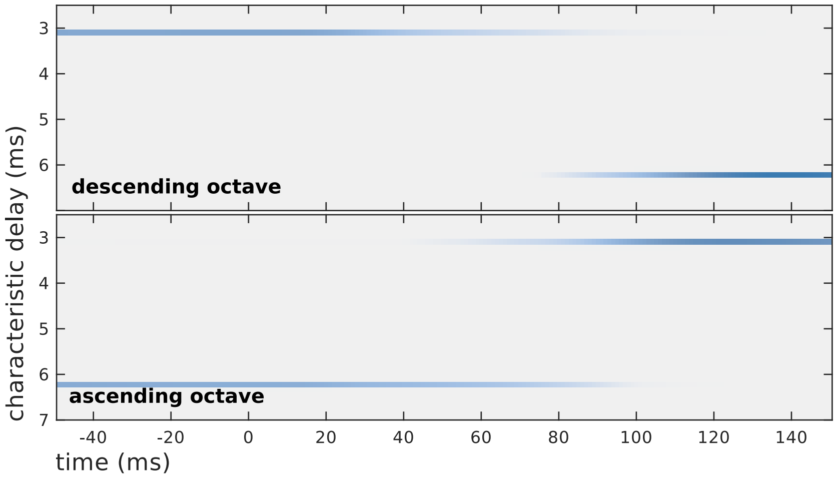


Figure B.3: Transition dynamics of octave shifts. Heatmaps show the transition dynamics of the inhibitory ensembles at the decoder during a pitch change between $f_0 = 160\,\mathrm{Hz}$ and its first harmonic $f_1 = 2\,f_0$. A) Descending (f_1 to f_0) interval. B) Descending (f_0 to f_1) interval. Note the slower build up and the comparatively increased offset delay in the descending condition. Simulation methodology was the same as in Figure B.1.

Conclusion

Horizontal interactions can effect the onset and offset timings of the tones involved. Transitions between tones are generally reflected in faster responses of the cortical system than when transitioning from or to silence.

Transitions between harmonically related tones present an asymmetric behaviour, reflecting the asymmetry of the harmonic inhibition in the populations at the decoder (see Figures 4.2A and B): transitioning to a higher octave results in a faster onset and slower offset of the previous tone; transitioning to a lower octave results in a slower, less salient transition.

Subtle timing differences are observed between transitions to different tones. There is a weak correlation between the onset/offset delays and the consonance percept elicited by a dyad comprising the tones of the interval, but it seems to reflect the harmonic structure of the tones rather than fundamental aspects of the perception of the transitions.

The effects described in this section are based on our model predictions; empirical testing should be addressed in future work using both, behavioural and MEG approaches. If the differential response to intervals comprising tones with different harmonic relations is empirically confirmed, its potential influence on the processing of interval judgements should be consider carefully.

Common abbreviations

A1/A2	primary/secondary auditory cortex
AC	auditory cortex
AEF	auditory evoked fields (MEG) (§2.3)
AEP	auditory evoked potentials (EEG) (§2.3)
AIM	auditory image model (§3.3.3)
alHG	anterolateral section of Heschl's gyrus (§2.1.3.1)
ALT HCT	alternated phase harmonic complex tones (§2.2.3.2)
AN	auditory nerve
BM	basilar membrane (§2.1.1.2)
CN	cochlear nucleus (§2.1.2.1)
CT	click train (§2.2.3.1)
DCM	dynamic causal model(ling) (§3.3.4)
E/I	excitation/inhibition
EIF	exponential (leaky) integrate and fire (§4.2.5.1)
EEG	electroencephalography
EOR	energy onset response (§2.3.2.4)
ESF	energy-related sustained field (§2.3.2.6)
FFR	frequency-following response (§2.3.2.1)
fMRI	functional magnetic resonance imaging
GPM	generative pitch model (§3.3.2)
HCT	harmonic complex tone (§2.2.3.1)
HG	Heschl's gyrus (§2.1.3.1)
IC	inferior colliculus (§2.1.2.1)
IRN	iterated rippled noise (§2.2.3.3)
LIF	linear (leaky) integrate and fire (§4.2.5.1)
LFP	local field potentials
MEG	magnetoencephalography (§2.1.3.1)
MGB	medial geniculate body (§2.1.2.1)
PET	positron emission tomography
pmHG	posteromedial section of Heschl's gyrus (§2.1.3.1)
POR	pitch onset response (§2.3.2.4)
PSF	pitch-related sustained field (§2.3.2.6)
PT	pure tone (§2.2.3.1)
SACF	summary autocorrelation function (§3.2.2.2)
STDP	spike-time-dependent plasticity (§3.2.3.2)
STI	strobed temporal integration (§3.3.3)
SF	sustained field (§2.3.2.6)
VCN	ventral cochlear nucleus (§2.1.2.1)