Aus der Neurologischen Universitätsklinik Heidelberg
Ärztlicher Direktor: Prof. Dr. med. Wolfgang Wick

Neuromagnetic representation of musical timbre dimensions in human auditory cortex

Inauguraldissertation
zur Erlangung des Doctor scientiarum humanarum
an der
Medizinischen Fakultät Heidelberg
der
Ruprecht-Karls-Universität
vorgelegt von
Melanie Günther
aus
Speyer
2024

Dekan: Herr Prof. Dr. med. Michael Boutros
Doktorvater: Herr Priv.-Doz. Dr. phil. André Rupp

For my grandparents.

Contents

Abbreviations

Abbreviation	Description
AEF	auditory evoked field
AIM	auditory image model
AMMA	Advanced Measures of Music Audiation
ANOVA	Analysis of Variance
ANSI	American National Standards Institute
AP	action potential
B	bassoon
BCa	bias-corrected and accelerated bootstrap
BM	basilar membrane
C	clarinet
cont	(melodies with) contour change
dB	decibel
f_0	fundamental frequency
F1/F2	formant 1/formant 2
fam	(melodies with) instrument family change
FM	frequency-modulated
fMRI	functional magnetic resonance imaging
HG	Heschl's gyrus
Hz	Hertz
IRN	iterated rippled noise
K	bass clarinet
kHz	kilo Hertz
l/L	left
loc	(melodies with) sound location change
LUFS	Loudness Units Full Scale

Abbreviation	Description
MDS	multidimensional scaling
MEG	magnetoencephalography
MFCC	Mel Frequency Cepstrum Coefficient
MGB	medial geniculate body
MMN	mismatch negativity
$N1_{on}$	onset-N1
$N1_{cng}$	N1 response to first tone of the melody change-N1
$N1_{post}$	N1 response to fourth tone of the melody post-change-N1
NAP	N1 response to fifth tone of the melody neural activity pattern
O	oboe
$P2_{on}$	onset-P2
$P2_{cng}$	P2 response to first tone of the melody change-P2
$P2_{post}$	P2 response to fourth tone of the melody post-change-P2
PAC	P2 response to fifth tone of the melody primary auditory cortex
POR	pitch onset response
PP	planum polare
PT	planum temporale
r/R	right
reg	(melodies with) instrument register change
RIS	regular interval sound
RMS	root mean square
SAI	stabilised auditory image
SPL	sound pressure level
SQUID	superconducting quantum interference device
STG	superior temporal gyrus
STS	superior temporal sulcus

List of Figures

List of Tables

1 Introduction

Since prehistory, music is a central component of human societies and culture (Zatorre and Salimpoor, 2013). The exact reasons for this are still unknown, but it is supposed that music fulfils important social functions which already pertain for more than 40,000 years (Zatorre and Salimpoor, 2013). Researchers coincide that music facilitates the expression and regulation of emotions (Zatorre and Salimpoor, 2013), which might help to avoid conflicts and strengthen social cohesion (Boer and Abubakar, 2014). As a cultural component, music creates a sense of togetherness and belonging which in turn may have a positive effect on individuals' personal well-being (Boer and Abubakar, 2014; Bowling, 2023). Many people also like listening to music because it produces a feeling of pleasure (Zatorre and Salimpoor, 2013). Zatorre and Salimpoor found that music is actually perceived as a reward, and the feeling of pleasure results from the activation of dopaminergic brain areas. The results of their study showed that music was liked and perceived especially rewarding when it fulfilled the listeners' expectations and thereby increased the dopamine release (Zatorre and Salimpoor, 2013).

Music also acts a means of communication (Fitch, 2006; Bowling, 2023). Music does not only transfer information about mood and emotion (Boer and Abubakar, 2014; Zatorre and Salimpoor, 2013; Bowling, 2023), but also about the identity of the instrument or speaker: the pitch and timbre of a sound reveal whether the instrument is large or small and the pitch and timbre of a vowel indicate whether the speaker is male, female or a child (Patterson et al., 2010a; Andermann et al., 2017). Though different modalities of communication, human speech and music share a lot of common characteristics (Fitch, 2006), e.g. both have a similar set of physical sound features (e.g. frequency, amplitude) and perceptual sound attributes (e.g. pitch, loudness, timbre). In particular, the perceptual attributes pitch, loudness and timbre and their processing are of major interest in auditory research. While the neurophysiological processing of pitch and loudness have already been extensively studied, there are still many open questions regarding the physical bases and physiological processing of timbre. The herein presented study addresses the question where and how timbre is processed in the human auditory cortex.

1.1 What Is Timbre?

Timbre is a complex sound feature for which no universally accepted definition exists. One of the first timbre definitions was provided by the physicist and physician Hermann von Helmholtz who described timbre as an internal perceptual sound attribute which represents but is not equal to the external physical sound (Siedenburg and McAdams, 2017). The most popular modern timbre definition originates from the American National Standards Institute (ANSI) and states that timbre is "that attribute of auditory sensation which enables a listener to judge that two non-identical sounds, similarly presented and having the same loudness, pitch and duration, are dissimilar" (ANSI, 1960, pg. 45; Plomp, 1970). The ANSI timbre definition is frequently cited, but it is also often criticised (Siedenburg and McAdams, 2017). Most critics relate to the fact that the ANSI timbre definition does not describe what timbre *is* but what it *is not* (e.g. Bregman, 1990, pg. 92–93; Schouten, 1970; Patterson et al., 2010b). Therefore, several researchers question the usefulness of this definition (e.g. Bregman, 1990, pg. 92–93) and search for a better, more concise timbre description (e.g. Siedenburg and McAdams, 2017). Some researchers already provided ideas for a more useful timbre definition; for example, the musicologists Reuter and Siddiq and the cognitive scientists Siedenburg and McAdams described timbre on the basis of its already known features and functions: Timbre is influenced by many different factors, including perceptual sound attributes like pitch and loudness (e.g. Siddiq et al., 2018; Siedenburg and McAdams, 2017), articulation and playing technique (e.g. Siddiq et al., 2018), but also material, architecture and shape of the musical instrument, as well as the type of sound production (e.g. double-reed mouthpiece, single-reed mouthpiece) (Patterson et al., 2010b). However, it was found that perceptual sound attributes, articulation and playing technique influenced timbre more strongly than the musical instrument as such (Siddiq et al., 2018; Siedenburg and McAdams, 2017). Furthermore, it was shown that one and the same instrument may produce a large number of different timbres – therefore, *timbre* can and should not be used to describe the 'general' sound of a musical instrument (Siddiq et al., 2017; Siddiq et al., 2018; Siedenburg and McAdams, 2017). Two other important functions of timbre are that it acts as a quality and also allows for the identification of sound sources; it applies to single tones, but also to entire musical phrases and musical pieces (Siedenburg and McAdams, 2017).

In a nutshell, timbre is a perceptual sound attribute which is influenced by several physical and perceptual factors (*dimensions*); thus, it is also referred to as a *multidimensional* perceptual sound attribute. Timbre applies to music and language, its acts at different scales, serves as a quality and allows for the identification of music-related and speech-related sound sources.

1.2 Anatomy and Physiology of Hearing

To understand the processing of timbre, it is essential to know the basic anatomy of the auditory system and the general physiological processes that take place during hearing. The human auditory system consists of two major parts, namely the central auditory system and the peripheral auditory system. The peripheral auditory system consists of the outer, middle and inner ear; the central auditory system comprises the cochlear nuclei, superior olivary nuclei, lateral lemniscus, inferior colliculi, medial geniculate nuclei, and the auditory cortex. In the following sections, the anatomy and functions of the peripheral auditory system and the central auditory system will be outlined. The literature used for this chapter was a neuroscience textbook by Bear et al. (2016) (pg. 378–414).

Anatomy of the Human Auditory System

One of the first important structures along the auditory pathway is the pinna of the outer ear which gathers sound waves and directs them through the ear canal towards the middle ear. The middle ear includes the tympanic membrane and the three auditory ossicles malleus, incus and stapes which transfer incoming sound waves to the inner ear. In addition to the vestibular system, the inner ear contains the cochlea which is the place where the auditory signal is converted into a neuronal signal. The cochlea has a bony and helically-shaped architecture and is bordered by two flexible membranes referred to as oval window and round window. Regarding its interior, the cochlea comprises three fluid-filled canals with different ion concentrations: the two outer canals *scala tympani* and *scala vestibuli* are filled with K^+-poor and Na^{2+}-rich perilymphe, whereas the middle canal *scala media* is filled with K^+-rich and Na^{2+}-poor endolymphe. The scala tympani, scala vestibuli and scala media are separated from each other by two membranes referred to as Reissner's membrane and basilar membrane.

The basilar membrane (BM) has a narrow and stiff basal end and a wide and flexible apical end. Directly on the BM, a structure called Corti organ is situated which contains the auditory receptors referred to as hair cells. According to their location within the Corti organ, the hair cells are differentiated into inner hair cells

and outer hair cells. Both the inner hair cells and outer hair cells have stereocilia at their apical end, but only stereocilia of the outer hair cells terminate in the overlying tectorial membrane. The stereocilia of both hair cell types are arranged in bundles in which the individual stereocilia are connected to each other by thin filaments called tip links. These filaments are linked to K^+ channels which are located near the tip of the stereocilia, and they connect the K^+ channels of one stereocilium with the K^+ channels of the neighboured stereocilium. At their basal end, both the inner hair cells and outer hair cells contact neurons called spiral ganglion cells. Again, there are differences between the two hair cell types: while each inner hair cell is connected to several spiral ganglion cells, each outer hair cell is connected to only one spiral ganglion cell. Thus, neurons connected to inner hair cells only receive input from one hair cell, whereas neurons connected to outer hair cells receive input from several hair cells. The axons of all spiral ganglion cells together form the auditory nerve which enters the vestibulocochlear nerve (cranial nerve VIII) and projects to the dorsal and ventral cochlear nuclei of the medulla. From the ventral cochlear nuclei, neurons project to the superior olives which in turn send axons via the lateral lemniscus to the inferior colliculi of the midbrain. It is important to note that the neurons of the cochlear nuclei only receive input from the ipsilateral ear (i.e. the ear that is located on the same body side), whereas the superior olives and subsequent structures receive input from both ears. The dorsal auditory pathway, which starts from the dorsal cochlear nuclei, bypasses the superior olives but also terminates at the inferior colliculi; there, the ventral auditory pathway and the dorsal auditory pathway converge and project to the medial geniculate body (MGB) of the thalamus. Finally, neurons of the MGB send axons to the primary auditory cortex (PAC) where the auditory information is further processed (fig. 1).

The primary auditory cortex (Brodmann area 41) is located in the superior temporal gyrus (STG) of the temporal lobe. Within the STG, the primary auditory cortex occupies the caudal two thirds of the transverse temporal gyrus which is also called Heschl's gyrus (HG) (fig. 3a). Two further regions named planum polare (PP) and planum temporale (PT) connect to the HG at its anterior and posterior ends, respectively. The highest instance of the auditory pathway is the auditory association cortex which consists of more widely distributed cortical areas which, however, are not subject of this work.

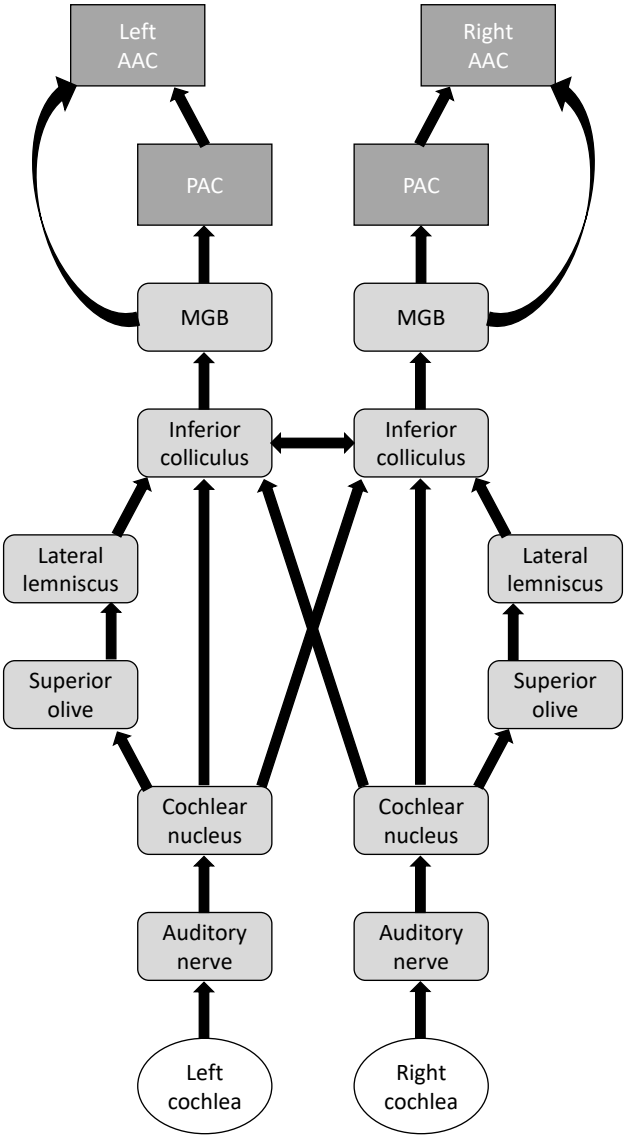


Figure 1: Schematic representation of information processing along the auditory pathway from the cochlea via the brainstem nuclei (light grey) to the auditory association cortex (AAC; dark grey). The graphic was created based on illustrations by Moore (2004) (pg. 50) and Andermann (2014) (pg. 10).

Basic Physiologic Processes along the Auditory Pathway

In response to incoming sound waves, the tympanic membrane starts to move and the motion energy is transferred via the auditory ossicles to the oval window. From there, the energy is transferred to the cochlear fluids and the basilar membrane. The initiated movement of the BM behaves like a wave; it starts at the BM's basis and continues towards the apex. Whether this movement eventually reaches the apex of the membrane or disappears earlier, depends on the frequency of the original auditory stimulus: stimuli with high frequencies (and high energy content) produce movements that lose most of their energy at the membrane's stiff basis and decay on their way towards the flexible apex; by contrast, stimuli with low frequencies (and low energy content) produce movements that barely lose energy at the membrane's basis and continue until the apex. As a consequence, stimuli with different frequencies also evoke maximal movements at different sites of the basilar membrane, leading to the emergence of a frequency place code along the basilar membrane.

The BM motion also affects the overlying Corti organ: when the basilar membrane moves, the stereocilia of the hair cells inside the Corti organ are either bent inwards (i.e. towards the modiolus) or outwards (i.e. away from the modiolus). While the tip links' tension increases and the K^+ channels further open if the stereocilia bend inwards, the tip links' tension decreases and the K^+ channels close if the stereocilia bend outwards. If the basilar membrane is motionless and the stereocilia are in neutral position, the tip links are moderately tense and the K^+ channels are partly opened. Only if stereocilia bend inwards, more K^+ ions flow into the respective hair cell and the hair cell depolarises. As a consequence of the depolarisation, voltage-gated Ca^{2+} channels in the hair cell's membrane open and release Ca^{2+} ions into the cytoplasm; the Ca^{2+} ions then provoke the release of neurotransmitter molecules which finally activate the post-synaptic ganglion cell(s). What kind of information is passed on to the spiral ganglion cells depends on the pre-synaptic hair cell(s): while the inner hair cells pass on most of the auditory information, the outer hair cells amplify the basilar membrane's responses such that even (very) soft sounds can be heard. Once activated by neurotransmitters, the spiral ganglion cells generate action potentials (APs) which then are forwarded from the cochlea to the auditory cortex. An important feature of the spiral ganglion cells is that each neuron responds to another frequency range and is most sensitive to one, so-called characteristic frequency. This characteristic frequency is of major importance for the frequency processing along the auditory pathway: as spiral ganglion cells with high characteristic frequencies contact hair cells near the apical end of the basilar membrane, and neurons with low characteristic frequencies contact hair

cells near the membrane's apical end, the frequency-place-code of the basilar membrane is preserved along the auditory pathway in tonotopic maps. These maps first emerge at the level of the cochlear nuclei and are transferred via the subsequent structures to the auditory cortex. Tonotopic maps and other strategies are finally applied for the processing of specific sound features such as frequency or sound location. Sound frequency is processed using two different strategies: frequencies below approx. 4 kHz are analysed using phase-coupled action potentials (i.e. action potentials which are evoked in synchrony with a specific sound wave phase); frequencies above approx. 4 kHz are coded using tonotopic frequency maps. Similarly, the processing of sound location also relies on two different strategies: sounds with a frequency between approx. 20 and 2000 Hz are localized using the interaural time difference (i.e. the time a sound needs to reach the second ear after having reached the first ear), whereas sounds with a frequency between approx. 2000 and 20,000 Hz are localized using the interaural intensity difference (i.e. the intensity a sound loses on its way to the second ear due to the head's sound shadow).

The processing of frequency and sound location both involve different structures along the auditory pathway. Cochlear nucleus neurons are especially sensitive to lower frequencies and respond with phase-coupled action potentials. Neurons of the superior olive use the resulting temporal AP pattern to calculate the interaural time difference and to determine the location of the sound; in the case of higher frequency sounds, superior olive neurons calculate the interaural intensity difference and apply it for sound localisation. Following the processing of frequency and sound location in the cochlear nuclei and superior olives, the MGB contributes to the processing of more complex sound characteristics of e.g. vowels and instrumental tones.

The physiologic processes along the auditory pathway can be simulated using computer models like e.g. the auditory image model (AIM; Patterson et al., 1995). More specifically, the AIM comprises different modules and filters which reliably reflect the physiologic processes in distinct anatomic structures from the outer ear to the brainstem nuclei (Ritter et al., 2005). In this study, AIM was applied to simulate the basilar membrane motion and the resulting neuronal activity pattern at the collicular level (Patterson and Holdsworth, 1991). An important approach for this is the so-called hair cell model which simulates the neurotransmitter currents within and around the inner hair cells and replicates the high temporal resolution and phase-coupling of the auditory nerve (Meddis, 1988; Andermann, 2014, pg. 15–16). Although periodic sounds such as instrumental tones or vowels evoke an oscillatory activity in the auditory nerve, the perception of these sounds is more or less stable. In AIM, this is replicated by strobed temporal integration on the

neural activity pattern: this form of temporal integration allows for creating a time-stable representation of the neural activity pattern while keeping the phase-locking fine structure of individual (processing) channels (Patterson and Holdsworth, 1991; Patterson et al., 1992). The sum of the channel-wise time-stable activity representations is referred to as stabilised auditory image (SAI) which is thought to reflect the initial sound perception (Patterson et al., 1995).

1.3 Magnetoencephalography

Magnetoencephalography or MEG is a non-invasive and silent neuroimaging technique which – due to its high temporal resolution ($< 1\,\text{ms}$) and good spatial resolution ($< 1\,\text{cm}$) – is a well-suited method for investigating the (early) neurophysiologic processes along the auditory pathway (Poeppel and Hickok, 2015). This section summarises the physical and physiological bases of MEG and provides more detailed information on the neuromagnetic activity which is elicited by acoustic stimuli.

Physical and Physiological Bases

Magnetoencephalography basically relies on ion currents that flow within single neurons and across neuron assemblies (Baillet, 2011, pg. 4; Baillet, 2017). The flow of charged ions produces small electric currents that add up and build larger net primary currents. Both within single neurons and neuron assemblies, the primary currents flow from a source to a sink and thereby create two differently charged electric poles (Baillet, 2011, pg. 3–4). Due to the conductive properties of the brain tissue, the primary currents induce further, so-called secondary currents, and both current types generate magnetic fields that can eventually be detected and measured using MEG (Baillet, 2011, pg. 3–4; Baillet, 2017). Most of the magnetic fields that are detected by the MEG result from large primary currents flowing through thousands of parallel pyramidal cells in the auditory cortex (Baillet, 2011, pg. 3; Baillet, 2017). As the magnetic fields are very small and much weaker than the Earth's magnetic field, they can only be detected under superconducting conditions (Baillet, 2011, pg. 5–7). In MEG, this is accomplished by superconducting quantum interference devices (SQUIDs) which are arranged in a helmet and cooled down to a temperature near absolute zero using liquid helium. The MEG device with its highly sensitive SQUIDs is installed in a shielded room (fig. 2) which drastically reduces the electromagnetic noise from outside and thereby creates the correct experimental conditions. During the measurement, the helmet of the MEG device is positioned above the subject's head such that the SQUIDs cover most of the subject's skull. Each of the typically 100–300 SQUIDs acts as an

amplifier of the measured magnetic field and represents the recorded data as a sensor waveform (Poeppel and Hickok, 2015; fig. 3b).

Figure 2: MEG device in an electromagnetically shielded room. The picture was created and kindly provided by Dr. Martin Andermann (University Hospital Heidelberg).

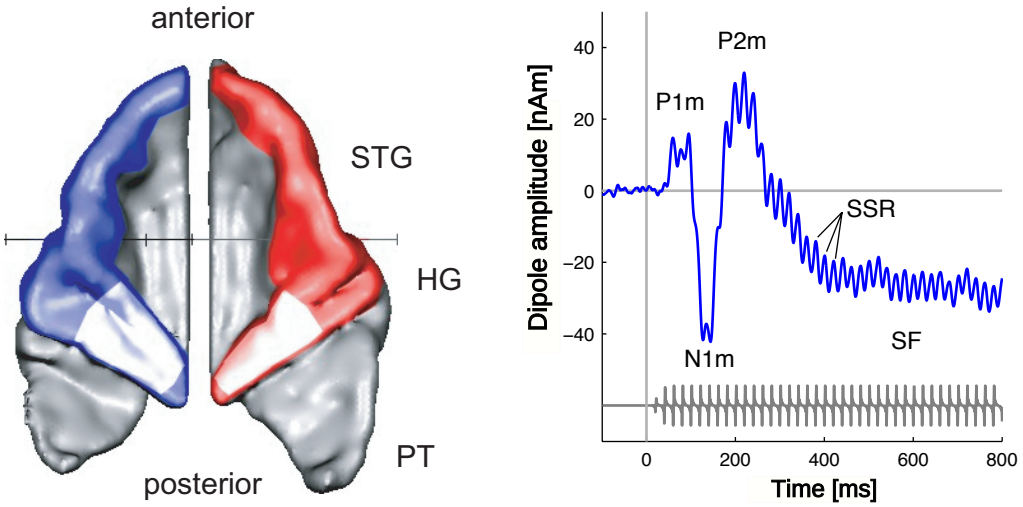
Neuromagnetic Responses

Neuromagnetic activity can be evoked by very different stimuli. The most common stimuli used in auditory research are tones, tone complexes, tone sequences, chords, short melodies, vowels, syllables, words and sentences (Poeppel and Hickok, 2015). All of these stimuli evoke neuromagnetic responses that consist of more or less the same components. There are four different types of components: transient responses, transition responses, auditory steady-state responses, and oscillations (Poeppel and Hickok, 2015). Transient responses are auditory evoked fields that occur in a specific sequence and near specific time points after the beginning of a sound (Poeppel and Hickok, 2015). Each sound with a minimum amount of energy and a minimum duration elicits three transient responses which are referred to as onset response, sustained field and offset response (Poeppel and Hickok, 2015). The onset response occurs directly at the beginning of a sound and marks the detection of sound energy; the sustained field begins after the onset response and persist as long as the sound is present, and the offset response marks the detection of the end of a sound (Poeppel and Hickok, 2015). Among the most important transient responses in auditory research are the P50m (P1), N100m (N1) and P200m (P2), which occur approximately 50 ms, 100 ms and 200 ms following sound onset and

have a positive (P1, P2) or negative (N1) sign. These transient responses are popular in auditory research because they can be found in every subject and their neural sources can be localized; thus, it is possible to find out where specific sound features are represented and processed in the auditory cortex (Poeppel and Hickok, 2015). Transition responses are auditory evoked fields that are not primarily evoked by sound energy but by changes of sound patterns (Poeppel and Hickok, 2015). Frequently investigated transition responses are the pitch onset response (POR), change responses, and mismatch negativity. Unlike the N1 which is elicited by the energy onset at the beginning of a sound, the POR is evoked by the pitch onset at the transition from a non-pitched sound to a pitched sound (Poeppel and Hickok, 2015). Especially large PORs are elicited by transitions from e.g. noise to a regular interval sound (RIS), as shown by Krumbholz (2003). In addition to the POR, there are also other change responses that occur at different kinds of sound pattern transitions. One very prominent transition response is the mismatch negativity (MMN) which is the amplitude difference between the response to a frequent 'standard' stimulus and the response to a rare 'deviant' stimulus. The usage of this so-called 'oddball' design is popular in auditory research because it is easy to implement and the MMN is a very sensitive and robust transition response (Poeppel and Hickok, 2015).

Source Locations

The source of a neuromagnetic response is its neural generator, i.e. a group of neurons that generates the measured activity. Finding the location of these sources is important in auditory research to find out where and how specific acoustic features are processed. This, however, is complex because the neural activity that was measured at the scalp theoretically may originate from any intracerebral location (inverse problem), i.e. there is an infinite number of possible source locations (e.g. Baillet, 2011, pg. 21–22). Due to the inverse problem, it is not possible to determine where a source is *definitely* located, but where it is *probably* located. The probability that the analysed source location corresponds to the actual source location, depends on data quality and on the amount of data that can be explained by the applied source model. Source models are typically created using source analysis software like e.g. BESA (BESA® GmbH, Gräfelfing, Germany) or MNE-Python (Python Software Foundation, MNE Developers, worldwide), which back-calculate source locations using a standardised three-dimensional head model, and output the coordinates of the most probable source location. Further information and details on source analysis and source modelling are given in the Methods section (section 2.1.5).
This work investigates the physical characteristics (amplitude, latency) and source locations of the transient responses N1 and P2 to find out how and where timbre is processed in the auditory cortex (fig. 3a).



(a) Auditory cortex **(b)** Auditory evoked field

Figure 3: (a) Axial view of the human auditory cortex. Lighter blue and red regions mark bilateral PAC, darker blue and red areas mark bilateral auditory association cortex. The three-dimensional representation was created by PD Dr. Peter Schneider (University Hospital Heidelberg) on the basis of magnetic resonance data and was provided by Dr. Martin Andermann (University Hospital Heidelberg). **(b)** Auditory evoked field measured in response to a complex harmonic sound (grey curve), obtained in the context of a previous study by Andermann (2014). The graph was created and kindly provided by Dr. Martin Andermann (University Hospital Heidelberg). Both graphs were previously published in Andermann (2014) (pg. 12).

1.4 Dimensions of Timbre

One of the most interesting but also most complex characteristics of timbre is its multidimensionality. In contrast to pitch and loudness research, timbre research began comparatively late. Only in 1863, when Helmholtz published his book "On the sensations of tone as a physiological basis for the study of music" (von Helmholtz, 1863; von Helmholtz, 1895), timbre research slowly began to evolve. A few decades later, the scientists Carl Stumpf (1890) and Gerhard Albersheim (1939) published their results on two further important timbre studies. While Helmholtz and Stumpf mainly focused on the physical aspects of timbre, Albersheim investigated the perceptual aspects of timbre (Wei et al., 2022). Albersheim was the first who proposed three perceptual timbre dimensions and displayed these dimensions in a so-called timbre space (Reuter, 2019). In this case, the timbre space was cylindrical and the three represented timbre dimensions were brightness, vowel type and saturation (Albersheim, 1939, pg. 353). The method of constructing timbre spaces

and using them for the graphical representation of the determined timbre dimensions became a new standard in timbre research (Reuter, 2019). Unfortunately, the timbre study by Albersheim was entirely theoretical and did not include confirmatory statistic calculations (Reuter, 2019). More advanced timbre studies were carried out in the later 20^{th} century by Schouten (1970), Grey (1977) and Wessel (1979). While Schouten and Wessel concentrated on the psychophysical aspects of timbre perception (Schouten, 1970; Wessel, 1979; Wei et al., 2022), Grey investigated the psychological aspects of timbre perception, by using several timbre dimensions (Grey, 1977; Wei et al., 2022). Grey's timbre study was the first empirical work on timbre dimensions and had a large impact on timbre research (Reuter, 2019). Still today, several timbre studies include the following methods:

1. the conduction of psychoacoustic experiments,
2. the usage of synthetic tones that are based on natural instrument tones,
3. the presentation of tone pairs that represent all (n-1)*n possible combinations of the n applied tones,
4. the subjects' task to rate the tones' (dis)similarity on a defined point scale,
5. the analysis of the (dis)similarity ratings using a multidimensional scaling (MDS) technique,
6. the graphical representation of the analysed timbre dimensions in a cubical timbre space.

Following Grey, especially Krumhansl and McAdams conducted timbre studies using multiple timbre dimensions and three- or four-dimensional timbre spaces (Reuter, 2019). The study by Grey yielded a three-dimensional cubical timbre space spanned by the dimensions (I) spectral energy distribution, (II) onset-offset pattern and (III) temporal patterns (Grey, 1977). Despite the fact that the study by Grey is still of great importance in timbre research, a drawback of his study was that he only used tones with the same pitch, loudness and duration (Reuter, 2019). The same can be observed in the timbre studies by Krumhansl and McAdams who applied an identical set of synthetic frequency-modulated (FM) sounds for their psychoacoustic experiments (Krumhansl, 1989; McAdams et al., 1995; McAdams, 1999). In addition to psychoacoustic measurements, McAdams et al. included questionnaires on the subjects' musical training and listening habits; on the basis of the subjects' answers, the authors classified the subjects into non-musicians, amateur musicians and professional musicians (McAdams et al., 1995; McAdams, 1999). Both the usage of questionnaires and the classification of subjects into categories became a common practice in timbre research (Reuter, 2019). Although Krumhansl and McAdams et al. used the same set of acoustic stimuli, they yielded different results: Krumhansl identified the timbre dimensions (I) spectral flux, (II) attack quality and (III) brightness (Krumhansl, 1989), McAdams et al. found the timbre dimensions

(I) attack time, (II) spectral centroid and (III) spectral flux (McAdams et al., 1995; McAdams, 1999). In order to investigate the commonalities and differences of the timbre spaces created by Grey, Krumhansl and McAdams, Siddiq et al. constructed a meta timbre space using tones from the three timbre studies (Siddiq and Reuter, 2013; Siddiq et al., 2014; Siddiq et al., 2015). Additionally, Siddiq et al. included tones from the Vienna Symphonic Library (Vienna Symphonic Library GmbH, Vienna, Austria) to compare the timbre of natural instrumental tones with the timbre of the synthesized tones applied in the three timbre studies. Data analysis, which was performed using MDS and agglomerative cluster analysis, revealed that tones originating from the same timbre study were more similar in timbre than tones originating from the same instrument (Siddiq et al., 2015; Reuter, 2019). Hence, Siddiq et al. concluded that the timbre of the applied tones and their position within the timbre space depended more on the tone set than on the respective instrument (Siddiq et al., 2015; Reuter, 2019). Furthermore, the results showed that timbre spaces differed a lot between studies and that it is therefore difficult to compare or even generalise timbre spaces (Reuter, 2019). A striking commonality of the four timbre studies by Grey, Krumhansl, McAdams and Siddiq is that all timbre spaces were created on the basis of tones with the same pitch, loudness and duration. In order to investigate the timbre of tones with different pitches and dynamic levels, Siddiq et al. carried out a timbre study in which they used natural tones from five instruments that were members of different instrument classes, and included three pitches and three dynamic levels (Siddiq et al., 2017; Siddiq et al., 2018); in sum, they used nine tones per instrument. Siddiq et al. found that timbre may vary a lot between tones of the same instrument and between tones of different instruments (Siddiq et al., 2017; Siddiq et al., 2018). Furthermore, they observed that the timbres did not form instrument-specific clusters; instead, the timbres of tones from different instruments largely overlapped (Siddiq et al., 2017; Siddiq et al., 2018). The obtained timbre space was spanned by four dimensions: (I) pitch, (II) brightness/spectral dynamic structure, (III) attack time and (IV) accompanying sounds (Siddiq et al., 2017; Siddiq et al., 2018; Reuter, 2019). Siddiq et al. found statistical proof for the first and second timbre dimensions, but not for the third and fourth timbre dimensions (Siddiq et al., 2017; Siddiq et al., 2018). Although timbre spaces may be very useful and practical – especially for displaying the different timbre dimensions – , their lack of standardisation and their dependence on the applied tone set may also make them rather useless e.g. in meta studies that use tones from several studies. In other cases, timbre spaces may be useful but not sufficient for addressing the research question. Therefore, in these or similar cases, researchers (additionally) apply other approaches (e.g. Mel Frequency Cepstrum Coefficients (MFCCs), formants) to investigate the dimensions of timbre (Reuter, 2019). Simply put, MFCCs (Davis and Mermelstein,

1980) are spectra of sound spectra over short time frames (Logan, 2000) and they can be used for the recognition and categorization of music instruments (Loughran et al., 2008), as well as for the calculation of sound (dis)similarity (Reuter, 2019). However, a disadvantage of MFCCs is that they are a lot less intuitive than timbre spaces (Reuter, 2019). Formants are fixed sound spectrum areas that have very large amplitudes in some partials (Schumann, 1929, pg. 89); they are independent from pitch and combine the advantages of MFCCs (i.e. standardised method, independence from tone set) and timbre spaces (i.e. intuitiveness) (Reuter, 2019). The disadvantage of formants is that they can only be used in timbre studies that include tones of wind or brass instruments which have more pronounced formant structures than string instruments (Reuter, 2019). An interesting timbre study that included both formants and MFCCs was carried out by Reuter and his colleagues: they compared the formants F1 and F2 of different orchestral wind instruments using a two-dimensional formant field spanned by the F1 and F2 frequencies, and conducted a psychoacoustic experiment using tones with different formant positions and formant distances (Reuter et al., 2017). In language, formant position and formant relation convey information about the speaker size and vowel type, respectively (Fitch, 1997; Patterson et al., 2008). In music, formant position characterises the instrument register and formant relation conveys information about the instrument family; thus, formant position and relation are important for differentiating and identifying musical instruments (Town and Bizley, 2013). In their study from 2017, Reuter et al. (2017) correlated the subjects' (dis)similarity ratings with the formant distances and the MFCCs: they found strong significant correlations between the (dis)similarity ratings and formant distances, and between the (dis)similarity ratings and MFCCs 1 and 2 (Reuter et al., 2018b; Reuter et al., 2018a). Furthermore, they showed that both formants F1 and F2 and MFCCs 1 and 2 can be applied for predicting timbre (dis)similarity (Reuter et al., 2018b; Reuter et al., 2018a). Based on these findings, this study primarily concentrates on the role of pitch, brightness, formant position and formant relation in timbre processing.

1.4.1 Pitch

Pitch is defined as the perceptual attribute of sound periodicity or as "that attribute of auditory sensation in terms of which sounds may be ordered on a musical scale" (ANSI, 1960, quoted from Plack and Oxenham (2005), pg. 1). Pitch is involved in both music and language and allows for identifying speakers (Gelfer and Mikos, 2005; D. Smith et al., 2005) and their emotional state (Fuller et al., 1992). Human pitch perception is restrained to periodicities between approximately 30 and 5000 Hz and the perceived pitch rises with increasing periodicity (Krumbholz et al., 2000). In many periodic sounds, pitch corresponds to the fundamental frequency f_0 – which is the common divisor of all sound's harmonics – and it was found that the auditory system uses both temporal and spectral cues for pitch extraction (Bendor et al., 2012). Therefore, pitch is not only the perceptual attribute of a temporal sound property (periodicity), but it can also be regarded a perceptual attribute of spectral sound properties. In this vein, it should also be mentioned that listeners may have different preferences for pitch perception: some mainly use temporal cues for pitch extraction (f_0 listeners), whereas others rather use spectral cues (overtone listeners) or do not have any preference for pitch extraction (neutral listeners) (refer to section 2.4.3 for more information).

Neural Processing of Pitch

The pitch of a sound is determined by using both the timing of phase-locked action potentials (Cariani, 1999) (up to frequencies of ca. 5000 Hz) and tonotopic frequency maps along the auditory pathway (Cohen et al., 1995). It is believed that sound periodicity is represented both at the subcortical level (e.g. Bidelman et al., 2010) and at the cortical level (e.g. Bizley and Walker, 2010). Especially the latter was confirmed by studies in which cortical lesions impaired the animals' ability to judge pitch changes of missing f_0 sounds (Whitfield, 1980). Similar impairments were found in humans with damages to the auditory cortex: subjects with bilaterally damaged HG had difficulties in differentiating pitches of harmonic complex tones (Zatorre, 1988; Tramo et al., 2002; Warrier and Zatorre, 2004) or frequencies of pure tones (Kazui, 1990; Tramo et al., 2002). Subjects with damages to the right auditory cortex were affected particularly frequently by difficulties in discriminating pitch (Sidtis and Volpe, 1988; Robin et al., 1990). Due to these findings, it is assumed that right HG may be involved in detecting the direction of pitch shifts (Johnsrude et al., 2000) and in determining the pitch of sounds with missing f_0 (Zatorre, 1988).

Several studies that investigated the processing of pitch in the human auditory cortex found that pitch is primarily processed in the anterolateral portion of the Heschl's gyrus (HG). One of these studies was carried out by Gutschalk et al. (2002) who investigated the perception of periodic sounds using regular and irregular click trains in an MEG experiment. Gutschalk et al. found that the lateral HG was especially responsive to regular sounds (i.e. pitch) and rather unresponsive to sound level; in contrast to this, the planum temporale (PT) was particularly sensitive to sound level and insensitive to pitch. These findings were corroborated and extended by Patterson et al. (2002) who conducted an fMRI experiment to study the hierarchy of pitch and melody processing. Patterson and his colleagues used three different kinds of acoustic stimuli: sounds without pitch, sounds with (fixed) pitch, and melodies. All three stimulus types activated the HG and the adjacent PT; however, differences were observed between sounds with pitch and sounds without pitch, and between sounds with pitch and melodies. Sounds with pitch caused stronger activation of the lateral HG than sounds without pitch, and melodies activated a larger area than sounds with pitch. The activation produced by melodies even extended to regions outside the HG and PT, mainly including the superior temporal gyrus (STG) and planum polare (PP). Furthermore, Patterson et al. (2002) found that both sounds without pitch and sounds with pitch produced a symmetric activation in bilateral HG; melodies, by contrast, caused an asymmetric activation with more activity in the right STG/PP than in the left STG/PP. This led to the conclusion that pitch is primarily processed in the lateral HG, while the processing of melodies probably occurs in the STG and/or PP. Another important fMRI experiment was performed by Penagos et al. (2004) who investigated the processing of pitch salience (i.e. the perceptual pitch strength). They presented harmonic complex tones with identical temporal regularity but differing pitch salience and found that all stimuli equally activated the cochlear nucleus, the inferior colliculus and the primary auditory cortex. However, an additional activity which was found in a non-primary auditory region overlapping lateral HG, varied with the pitch salience of the presented stimuli: sounds with a high pitch salience produced a strong activation, while sounds with a low pitch salience produced a weak activation. Thus, the results confirmed that the HG is involved in pitch processing, and further revealed that pitch salience is represented in lateral HG. The HG is also the place where the POR is elicited in response to pitch changes during continuous iterated rippled noise (IRN) sounds (i.e. noise sounds that are created by iterative copying and time-delayed adding (Yost, 1996). More specifically, the POR is evoked in lateral HG when there is a pitch change from one IRN sound to another (Ritter et al., 2005), and it is evoked in medial HG when a periodic sound is presented following noise (Krumbholz, 2003). It was shown that the latency and amplitude

of the POR are influenced by pitch height and pitch salience (Krumbholz, 2003; Ritter et al., 2005). The neuromagnetic activity elicited at pitch onset and pitch offset was found to mirror the absolute f_0, while the activity evoked by pitch changes within melodic sequences reflected the relative f_0 (Andermann et al., 2021). The amplitude of the neuromagnetic responses was smaller for sequences with fixed f_0 and larger for sequences with variable f_0. This effect was significantly stronger in musicians than in non-musicians, and musicians showed larger neuromagnetic responses in general. Consequently, it was assumed that the cortical processing of pitch is modulated by musical experience (Andermann et al., 2021). Supporting findings were observed in a fMRI study by Bianchi et al. (2017): musicians showed better performance in pitch discrimination tasks and were more sensitive to small pitch changes than non-musicians. Furthermore, the musicians' sensitivity to pitch changes was correlated with the evoked neural activity in the right auditory cortex; this was not observed in non-musicians. However, across all subjects, a significant correlation was found between the pitch discrimination thresholds and the evoked activity at the inferior colliculi. It was therefore suggested that musical training influences pitch processing from the subcortical level (inferior colliculi) up to the cortical level (auditory cortex).

All in all, the processing and neural representation of pitch have been studied extensively and there are a lot of further findings on this topic which, however, would go far beyond the scope of this work. Nevertheless, it should be noted that – although the HG indeed is a pitch processing hub – there is increasing evidence that pitch may be processed in a network comprising also widely distributed cortical areas. For instance, it was found that pitch chroma is processed anterior to primary auditory cortex (Warren et al., 2003) and pitch patterns of melodic sequences are processed in more distributed frontoparietal regions with increasing lateralisation (Zatorre et al., 1994; Patterson et al., 2002). Finally, data from lesion studies support the view that pitch is processed in distributed but locally specialized cortical regions (Stewart et al., 2006).

Pitch as Timbre Dimension

Pitch has not always been considered a timbre dimension (e.g. Marozeau et al., 2003) and until now, there are discussions on whether pitch really is a timbre dimension or not. However, researchers largely coincide that timbre and pitch are interdependent and greatly influence each other (e.g. Marozeau and de Cheveigné, 2007; Allen and Oxenham, 2014; Melara and Marks, 1990; Krumhansl and Iverson, 1992). Part of this interaction between timbre and pitch can be explained on a physical basis: high pitches usually come with increased brightness because the spectrum of high-pitched sounds contains many pronounced high partial tones;

similarly, decreased brightness is often linked to low pitches because the spectrum of low-pitched sounds contains fewer pronounced high partials (Siedenburg and McAdams, 2017). Another important finding is that changes in timbre affect the participants' ability to determine minor changes in pitch (Allen and Oxenham, 2014). Subjects were shown to have more difficulties in concentrating on pitch changes while ignoring changes in timbre, and less difficulties in concentrating on timbre changes while ignoring changes in pitch (Beal, 1985). Beal (1985) and Pitt (1994) investigated whether pitch and timbre changes can equally be discriminated by musicians and non-musicians. They found a better performance in musicians, but both studies lacked a statistically relevant number of stimuli and perceptual salience was not controlled for the applied pitch and timbre changes (Allen and Oxenham, 2014). Melara and Marks (1990) and Krumhansl and Iverson, 1992 also found interactions between the perception of pitch and the perception of timbre, but similar to Beal (1985) and Pitt (1994), the number of the applied stimuli was too small for reaching statistical relevance (Allen and Oxenham, 2014).

1.4.2 Brightness

Brightness is a perceptual sound attribute and describes how 'bright' or 'dull' a tone sounds. Brightness is strongly associated with a sound's spectral energy distribution (Grey and Gordon, 1978; Wessel, 1979) and in particular with its spectral centroid (Peeters, 2004; Lartillot et al., 2008), i.e. the spectrum's center of gravity. In timbre studies, the spectral centroid is a popular tool for extracting the brightness of a sound. The rationale behind this is that a higher center of gravity is linked to increased energy in higher frequency ranges which leads to the perception of more brightness (Czedik-Eysenberg, 2016, pg. 34). Consequently, sounds with a higher spectral centroid are perceived 'brighter', while sounds with a lower spectral centroid are perceived 'darker' (Caclin et al., 2005). A disadvantage of this interaction is that the spectral centroid is also linked to the fundamental frequency: as the harmonics of a sound increase with rising f_0, more energy is present at higher frequencies and the spectral centroid rises as well (Czedik-Eysenberg, 2016, pg. 34). Therefore, sounds with higher pitches are commonly perceived brighter than sounds with lower pitches (Czedik-Eysenberg, 2016, pg. 34). In contrast to pitch, there seems to be consensus that brightness is a dimension of timbre. In fact, several studies showed that the spectral centroid is a salient timbre descriptor and can be applied for dissimilarity ratings (Caclin et al., 2005). Like timbre in general, brightness or the spectral centroid were also found to interact with pitch. For example, a study by Allen and Oxenham (2014) investigated interference effects between changes in spectral centroid (which elicited changes in brightness) and changes in f_0 (which elicited changes in pitch) and tested whether there are differences between musicians (subjects with at least

eight years of musical training) and non-musicians (subjects with two or less years of musical training). In line with the findings of Beal (1985) and Pitt (1994), they found lower discrimination thresholds for f_0 differences in musicians, but the discrimination thresholds for spectral centroid changes did not differ significantly between musicians and non-musicians. Interaction effects were found for both types of stimuli: the subjects' ability to determine f_0 changes was impaired by changes in spectral centroid, and their ability to determine spectral centroid changes was impaired by changes in f_0. This finding applied to both musicians and non-musicians and all subjects were similarly susceptible to confusions between changes in f_0 and changes in spectral centroid. Due to these findings, Allen and Oxenham (2014) supposed that pitch and timbre are processed symmetrically and that the susceptibility to interaction effects is not linked to the amount of musical training. In a subsequent study, Allen et al. (2016) investigated the representation of pitch changes and timbre changes in the human auditory cortex. Again, they used stimuli with changes in f_0 to elicit changes in pitch and stimuli with changes in spectral centroid to evoke changes in brightness. Allen et al., 2016 observed that sequences with more f_0 or spectral centroid variation evoked a stronger activity in the auditory cortex than sequences with less variation. They also found that sequences with f_0 changes and sequences with spectral centroid changes activated largely overlapping cortical areas. However, within the individual subjects, Allen et al. (2016) observed different activation patterns between the two stimulus types; thus, pitch and timbre might be processed separately within the same regions of bilateral auditory cortex. Allen and her colleagues refined their findings in a later publication on pitch and timbre processing: pitch and timbre were represented in largely overlapping cortical areas but the attention to either f_0 changes or spectral centroid changes altered the activation patterns even if the stimuli did not differ physically (Allen et al., 2019).

Analogous to spectral centroid in music, formants influence brightness in speech. Studies investigating the processing of vowel type and speaker size in human auditory cortex revealed that changes in formant position and formant relation both activated anterior PT, while changes in pitch primarily activated posterior HG (Andermann et al., 2017; von Kriegstein et al., 2006; von Kriegstein et al., 2007; von Kriegstein et al., 2010). These results were also in line with findings from previous fMRI and MEG studies on the processing of temporal and spectral sound features (e.g. Andermann et al., 2011; Griffiths et al., 1998; Patterson et al., 2002; Gutschalk et al., 2004; Ritter et al., 2005) and indicated that HG is primarily involved in temporal processing, while PT is mainly implicated in spectral processing and the extraction of formant position and formant relation (Andermann et al., 2017). Although the neural sources of temporal and spectral processing were

located close to each other, there was a clear functional separation: neuromagnetic responses to changes in pitch, formant position and formant relation formed part of the early N1 component which indicated that these sound features are processed early in the auditory hierarchy (Irino and Patterson, 2002; Patterson and Irino, 2014), and that there is an early segregation of content-related information ('what?') and speaker information ('who?') (Formisano et al., 2008; Nelken et al., 2014).

1.4.3 Attack Time

The attack time corresponds to the time interval from a sound's energy onset to steady state, i.e. from the beginning of the temporal envelope to its maximal amplitude (Bizley and Walker, 2010). In music instruments, the duration and form of the sound onset largely influence the perceived quality and allow for an initial identification of the instrument (Bizley and Walker, 2010). Depending on the amount of melodic and percussive elements, the onset of a sound may be rather smooth or abrupt which eventually influences the perceived timbre (e.g. Iverson and Krumhansl, 1993). Similar to the spectral centroid, there is little doubt that attack time is a dimension of timbre. Results from earlier scaling studies already indicated that dynamic aspects might be involved in timbre perception. In a study on the role of attack time in the context of timbre, Wedin and Goude (1972) removed the onsets of natural tones and presented both the complete sounds and incomplete sounds in a similarity rating experiment. The subjects' ratings did not differ much between the complete sounds and the incomplete sounds, but still the results indicated that dynamic aspects influenced the similarity judgements. Further studies eventually led to the identification of the involved temporal sound properties. For instance, it was found that tones without low-amplitude energy at their onset sounded less natural than tones with unchanged onset (Grey and Moorer, 1977) and that subjects' similarity ratings correlated with characteristics of the sound onset (Wessel, 1979; Krumhansl, 1989). Studies on the identification of musical instruments corroborated these findings: Clark et al. (1963) showed that subjects performed poor when the attack was removed from sounds, but they performed well when the attack was present and also when they only heard the first 60 ms of the sounds. A similar finding was obtained by Saldanha and Corso (1964): he observed that the removal of the attack largely impaired sound recognition and that the combination of attack plus steady state was necessary for correct identification. Although some of these findings could not be confirmed by other authors (e.g. Iverson and Krumhansl, 1993), there is increasing evidence that the attack of a sound may actually be a major timbre descriptor. Suied et al. (2014) showed that listeners could categorise sung voices, percussive sounds and string instrument sounds already 4 ms (voice) and 8 ms (instruments) after sound onset. A further study carried out by Ogg et al. (2017) revealed that subjects could

differentiate instrumental sounds, human speech sounds and human environmental sounds robustly after approximately 25 ms; moreover, they found that the presence of an attack improved the discrimination of both speech and instrumental sounds. According to Siedenburg (2019), the presence of an attack is not required for the correct identification of musical instruments. But if it is not required, why do listeners at all use the attack time for sound and instrument identification? This question cannot be answered to this day – it is still unclear why the attack of a sound is rather used for this purpose than other sound parts. Researchers suppose that the attack contains more information on the sound source than sustained sound parts (Siedenburg, 2019); however, this applies mainly to percussive instrumental sounds (Suied et al., 2014) which hardly contain any steady state (Campbell and Greated, 1994, pg. 13). Regarding non-percussive instrumental sounds and speech sounds (vowels), the sustained sound parts seem to be equally or even more useful for the identification of sound sources than the attack time (Suied et al., 2014).

1.4.4 Sound Location

Being able to localise sound sources is of major importance for both humans' and animals' survival. Sound localisation begins relatively early in the auditory pathway and involves several brainstem nuclei and the midbrain; however, it was shown that the auditory cortex also has a major role in sound localisation (Bizley and Walker, 2010). While the subcortical regions use interaural time and intensity differences to localise sound sources, the auditory cortex seems to combine the different location cues to an auditory space percept (Bizley and Walker, 2010). In fact, results from lesion studies indicated that the auditory cortex uses both interaural time/intensity differences and monaural spectral cues (Bizley and Walker, 2010) to localise sound sources in the horizontal plane (A. Smith et al., 2004) and in the vertical plane (Bizley et al., 2007). Previously, it was found that subjects with unilateral lesions to the auditory cortex were impaired on contralateral sound localisation (Kavanagh and Kelly, 1987; Malhotra and Lomber, 2007) and that the extent of the impairments correlated with the size of the lesion (Bizley et al., 2007). Further results from lesion studies implied that the auditory cortex might contain specific pathways for sound localisation (Clarke et al., 1996; Clarke et al., 2000), but researchers are still divided as to whether spatial sensitivity is centralised to distinct cortical networks or whether it is distributed throughout the auditory cortex (Bizley and Walker, 2010). In humans, the planum temporale was shown to be especially sensitive to spatial cues and it is regarded an important center of spatial processing (e.g. Zatorre and Penhune, 2001; Griffiths and Green, 1999). Auditory cortical regions that are particularly sensitive to spatial cues were also identified in cats and primates; however, in both animal groups, neurons with

spatial sensitivity were found throughout the auditory cortex (Bizley and Walker, 2010). Consequently, more investigation is needed to clarify where and how sound location is processed in the (human) auditory cortex.

1.5 Neural Processing of Timbre

Several findings on the processing of timbre were obtained from lesion studies which were performed on patients with temporal lobe damages. One important observation was that the ability to detect timbre changes was particularly impaired in patients with right temporal lobe lesion but intact left temporal lobe (Samson and Zatorre, 1988). Patients with right temporal lobe lesions were not only impaired in detecting changes in temporal timbre (attack time), but also in detecting changes in spectral timbre (complex tones with different harmonics) (Bizley and Walker, 2010); by contrast, these impairments were not observed in patients with lesions of the left temporal lobe (Samson and Zatorre, 1994). The results of later studies implied that both the right and left temporal lobes are involved in timbre processing: patients with left temporal lobe lesions did not have difficulties in discriminating single tones with different onset characteristics, but they could not differentiate the tones when these were presented in a melody; patients with right temporal lobe lesions were not able to differentiate the tones, neither in the single-tone mode, nor in the melody mode (Samson et al., 2002). Further findings regarding the neural substrate of timbre processing were obtained – among others – from electrophysiological studies: Warren et al. (2005) showed that several overlapping areas of the auditory cortex are involved in lower-level spectral analysis (i.e. processing of pitch and spectral envelope), whereas an area in the midportion of the right superior temporal sulcus (STS) is implicated in higher-level spectral analysis (i.e. processing of fine frequency spectrum) which in turn allows for timbre differentiation. Thus, it is believed that the processing of timbre may be lateralised to the right hemisphere and that there is a hierarchy in spectral processing. This assumption is supported by the finding that the spectral envelope is processed in a line from the HG via the PT to the STG (Kumar et al., 2007). Although the number of studies is rare, there are indications that a sound's spectrum is indeed analysed at several spectral and temporal scales (Shamma et al., 1993). Studies in animals revealed that the auditory cortex contains neurons that are differentially sensitive to temporal sound properties (e.g. amplitude envelope, repetition rate) (Lu et al., 2001; Sakai et al., 2009), sound intensity (Heil et al., 1994) and sound onset characteristics (Heil, 1997a; Heil, 1997b). The differential sensitivity of neurons to temporal and spectral sound properties might eventually be the reason why some neuron clusters in the primary auditory cortex are differentially sensitive to e.g. speech or location cues (Wallace and Palmer, 2009). Similar results were obtained in studies on vowel

processing. For instance, it was found that higher cortical areas including the STG are sensitive to vowel class (Obleser et al., 2005) and that vowel space is represented in cortical maps on the basis of acoustic features (Scharinger et al., 2011). Topographic representations of acoustic features were not only found in humans, but also in different animal species (e.g. Bizley et al., 2009; Bizley and Walker, 2010; Bendor and Wang, 2005; Bendor and Wang, 2006). Although the findings of animal studies are not directly transferable to humans, they still allow for a better understanding of timbre processing in general. For instance, one animal study showed that the primary auditory cortex of gerbils contains maps of formant distances even at the cellular level (Ohl and Scheich, 1997). A human fMRI study revealed that vowel and speaker identity were mainly processed in bilateral STG, anterior-lateral HG, PT, and in STS; however, the processing of speaker identity was more pronounced in the right hemisphere and interspersed with vowel sensitive regions (Formisano et al., 2008). Therefore, these findings support the hypotheses that timbre perception is accomplished by the interplay of several distributed cortical areas and that low- and high-level acoustic features (like vowel identity and speaker identity, respectively) are processed in a hierarchical way (Town and Bizley, 2013). A study that investigated the representation of three sound categories (guitar, cat, singer) in auditory cortex showed that the sounds were assigned according to their timbre and that the sounds of different categories elicited different spatial activation patterns in auditory cortex; however, all sounds activated the anterior-lateral HG, PT, posterior STG and/or STS (Staeren et al., 2009). Similar observations were also made regarding the processing of pitch (section 1.4.1). Researchers believe that sound identity and acoustic scale are encoded in the spectral envelope of sounds (van Dinther and Patterson, 2006; von Kriegstein et al., 2006). This would explain why bilateral STG is activated by both changes in identity and changes in size; however, the left posterior STG is more activated by auditory scale changes in human voice, whereas the anterior temporal lobe and intraparietal sulcus are sensitive to size changes in any 'voice' (human, animal, instruments) (von Kriegstein et al., 2007). Although the right hemisphere was shown to have a special role in timbre processing and perception, there is evidence that the left hemisphere might be selectively activated by acoustic features to perform distinct functions (Deike et al., 2004; Town and Bizley, 2013). Researchers could also identify cortical regions that are especially involved in timbre processing, but the search for timbre-specific neurons remains difficult. In ferrets, the neurons of the auditory cortex were sensitive to timbre, but they were also sensitive to other sound attributes like pitch and location. Instead of being timbre-specific, the neurons of distinct cortical regions responded differentially to timbre, pitch and location cues, e.g. some neurons showed a preference for high pitches, while other neurons were more sensitive to specific vowels (Bizley et al., 2009). This observation is in line with the finding

that low- and high-level acoustic features are represented across multiple regions of human auditory cortex (Staeren et al., 2009). A big advantage of vowels in timbre research is that spoken and artificially generated vowels have large steady states and rather simple temporal structure; by contrast, natural and instrumental sounds have more complex spectral and temporal structures (Bizley and Cohen, 2013). Thus, considering both the spectral and temporal acoustic features is of major importance when investigating musical timbre (e.g. Samson and Zatorre, 1994; McAdams et al., 1995; Elliott et al., 2013).

1.6 Objectives of the Study

The current study was performed to investigate the neurophysiological processing and neuromagnetic representation of musical timbre dimensions in the human auditory cortex. More specifically, a combination of MEG and psychoacoustic measurements was used to extract the neural responses to changes in pitch contour, instrument register, instrument family, sound location and to find out whether and how the neural activity and the location of the neural sources correlate with the perceived brightness (which is a major timbre determinant) of the applied instrumental tones. Furthermore, the effect of natural and standardised attack times on timbre processing was investigated. The motives for this study were manifold: first and foremost, timbre is an important perceptual attribute in music and speech and although it plays a crucial role in human communication, little is known about its nature and how it is processed in the human brain. Secondly, the number of investigations dealing with timbre is relatively small when compared with other perceptual sound attributes such as pitch. The fact that pitch and timbre interact and that they can hardly be perceived independently, implies that more knowledge about timbre is needed to gain a better overall understanding of music and language processing. A secondary aim of this study was to find out whether and how musicality and musical experience impact timbre processing. Regarding pitch processing, significant correlations were found between the amplitude/latency of the neuromagnetic responses and the subjects' musical aptitude: subjects with higher musical aptitude responded more strongly and faster to pitch shifts than subjects with lower musical aptitude (Andermann et al., 2021). As pitch and timbre largely interact, the question arose whether musicality and/or musical experience might have a similar effect on timbre processing as on pitch processing. In order to find out whether and how musicality/musical experience influence timbre processing, a musicality test and a questionnaire on musical experience were applied. Furthermore, a pitch perception test was conducted to investigate whether the subjects' pitch perception type might affect timbre processing, as it was found for pitch processing (Schneider et al., 2005). In line with findings from previous

studies (e.g. Andermann et al., 2021), it was expected that melodies with changes in pitch contour/instrument register/instrument family/sound location elicit stronger and earlier N1 and P2 responses than melodies without these changes. Furthermore, it was assumed that melodies with contour change might evoke relatively weak responses, while melodies with register change might elicit relatively strong N1 and P2 responses. Melodies with standardised attack times were supposed to evoke stronger and earlier N1 and P2 responses than melodies with natural attack times because stimuli with more temporal regularity were shown to elicit stronger and earlier responses than stimuli with less temporal regularity (e.g. Kim et al., 2022). Consistent with findings from pitch studies (e.g. Andermann et al., 2021), it was assumed that listeners with higher musical aptitude elicit stronger and earlier neuromagnetic responses than listeners with lower musical aptitude. At the same time, no significant differences between fundamental pitch listeners and overtone listeners were observed in these studies; therefore, a similar result was expected in the herein study. In connection with the above mentioned hypotheses, the neural source locations were expected to differ between melodies with and without contour/instrument register/instrument family/sound location changes (right–left, left–right), and between listeners with higher and lower musical aptitude.

2 Materials and Methods

The methods applied in this study comprise an MEG experiment, a psychoacoustic timbre experiment and further psychoacoustic tests (musicality test, pitch perception test). Additionally, the subjects' hearing abilities were tested in an audiometric screening and the individual musical experience was assessed in a short questionnaire.

2.1 Magnetoencephalography Experiment

2.1.1 Participants

Thirty-four healthy subjects (19 female, 15 male, 27.9 ± 8.9 years, 2 left-handed) participated in the MEG experiment after providing written informed consent. None of the participants reported a current or past neurological or psychiatric disorder or hearing impairments. All 34 subjects participated in two consecutive MEG measurements which will be referred to as MEG runs A and B. The experimental procedures in this work were approved by the ethics board of the Medical Faculty (University of Heidelberg; numbers of approval: S-959/2021).

2.1.2 Stimuli

In both MEG runs, melodies played by an oboe, a bassoon, a clarinet, and a bass clarinet were presented. Each melody consisted of two tone triplets which were made up from the tones D2, F♯2 and A2 (low register) or D4, F♯4 and A4 (high register). The low register tones were played by the bassoon and the bass clarinet, whereas the high register tones were played by the oboe and the clarinet. According to the type of sound production, the double-reed instruments oboe and the bassoon were grouped to an *oboe family*, while the single-reed instruments clarinet and the bass clarinet were grouped to a *clarinet family* (fig. 4). Overall, there were five different types of tone triplets (fig. 5) which differed in the general tone sequence, irrespective of the register. Depending on the type of the two applied tone triplets, the resulting melody either had a contour change, a register change, a family change, or a sound location change at the transition point (tone 4) from the first triplet to the second triplet (fig. 6).

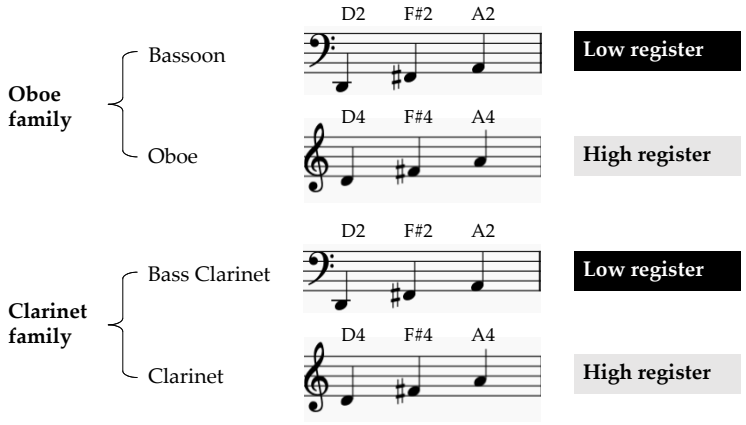


Figure 4: Tones that were used for the creation of tone triplets and melodies. Each three tones were played by an oboe, a clarinet, a bassoon or a bass clarinet. The oboe and the bassoon together formed an oboe family, while the clarinet and the bass clarinet formed a clarinet family. Simultaneously, the oboe and the clarinet built a group of high register instruments, while the bassoon and the bass clarinet formed a group of low register instruments.

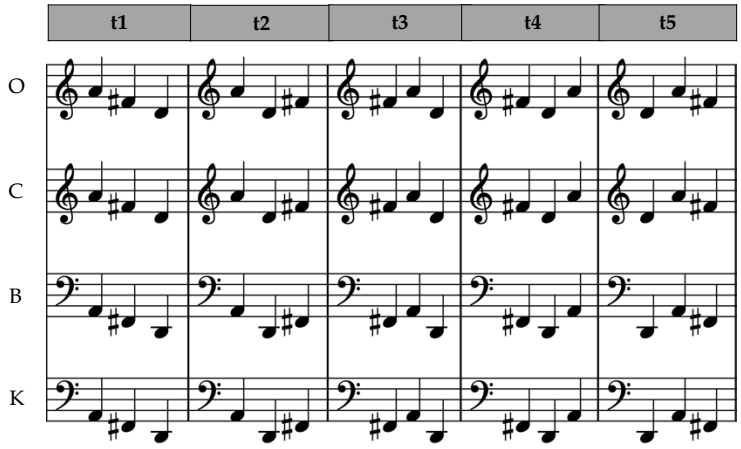


Figure 5: Tone triplets that were used for the creation of melodies. Five different types of tone triplets (t1–t5) were played by an oboe (O), a clarinet (C), a bassoon (B), or a bass clarinet (K). The general tone sequence of each triplet type was constant across the four instruments and the two registers.

All tones that were used for the creation of the tone triplets and melodies were obtained from the music notation software MuseScore 3.6 (MuseScore Ltd, Limassol, Cyprus). Using MATLAB R2022a (The MathWorks, Inc., Natick, USA), the tones were cut to a length of 600 ms, a sampling rate of 48,000 Hz and a bandwidth filter from 20 to 4000 Hz were applied, and loudness was normalized to -23 LUFS (EBU R 128 standard). The attack time was specified by Hanning windows whose length was either 2 ms to use the natural attack time (while avoiding clipping of the sound), or 75 ms to use an average-like attack time that can be applied to all deployed sounds; the offset of all sounds was set to 50 ms. Finally, loudness was attenuated unilaterally (left or right side) by -6 dB to induce the perception of an unilateral spatial sound location. At the end of this procedure, the total set of acoustic stimuli comprised 64 melodies which consisted of 2 x 32 melodies with the same type of attack time, or 4 x 16 melodies with the same type of change. In accordance with their type of attack time, the melodies were divided into the groups *melodies with natural attack times* and *melodies with unified attack times*. Melodies with the same type of change were divided into the groups *melodies with contour change*, *melodies with register change*, *melodies with family change* and *melodies with sound location change*. The melodies of these four melody groups can be characterised as follows:

1. **Melodies with contour change** consisted of two tone triplets that differed in tone sequence. The entire melody was played by the same instrument and sound location was fixed (table 25).

2. **Melodies with register change** consisted of two tone triplets that differed in register but not in the general tone sequence. One triplet was played by an instrument of the high register, whereas the other triplet was played by an instrument of the low register; the instrument family and sound location remained the same (table 26).

3. **Melodies with family change** consisted of two tone triplets that differed in family but not in the general tone sequence. One triplet was played by an instrument of the oboe family, whereas the other triplet was played by an instrument of the clarinet family; the register and sound location remained the same (table 27).

4. **Melodies with sound location change** consisted of two tone triplets that differed in sound location. One triplet had a sound location on the right side, whereas the other triplet had a sound location on the left side. The entire melody was played by the same instrument (table 28).

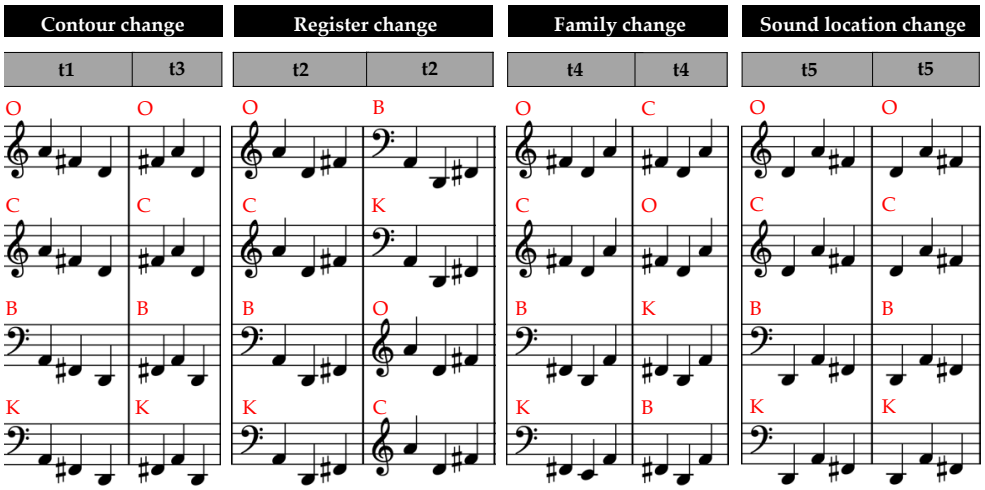


Figure 6: Melodies that were used for the creation of the acoustic stimuli. Each four melodies had a contour change, a register change, a family change, or a sound location change. Within each melody group, there was one melody each played by an oboe (O), a clarinet (C), a bassoon (B), or a bass clarinet (K). Every melody was built from two tone triplets (t1, t2, t3, t4, or t5) which were either identical or different. Within each melody group, the general tone sequence was the same across the four instruments and the two registers.

2.1.3 Paradigm

Two paradigms were created for the two MEG runs: the paradigm for MEG run A included the 32 melodies with natural attack times, and the paradigm for MEG run B included the 32 melodies with unified attack times. In each MEG run, each melody was presented 16 times, i.e. 512 melodies were presented per run. As the two paradigms had the same structure, the following information refers to both paradigms: within each of overall 16 rounds, the 32 melodies were presented in a random order, such that, in the course of the measurement, the melodies were presented in 16 different orders. The resulting sequence comprising all 512 melodies was used for both MEG runs A and B that were carried out on the same subject. For each of the 34 subjects, another melody sequence was created according to the same criteria, i.e. the applied melody sequence differed between the subjects. To avoid further possible sequence effects, the order in which the two MEG runs were conducted, was reverted for about half of the subjects. More precisely, 18 subjects first participated in MEG run A and then in MEG run B; 16 subjects first participated in MEG run B and then in MEG run A.

2.1.4 Measurement

Prior to the beginning of the MEG experiment, the participants were asked to change their clothes for metal-free clothes and to take off all other removable metal objects. Persons who had a retainer or magnetic implants were excluded from the study. If the latter was not the case and participation was possible, the subject was prepared for the MEG measurement: first, four coils were attached to the head at the left and right mastoids and the left and right corners of the forehead. Next, the shape of the subject' head was digitised using a Polhemus 3D-Space Isotrack2 system: a reference point was attached to the centre of the forehead and the relative positions of three anatomical landmarks (left and right pre-auricular points, nasion) and the four coils were measured; additionally, 100 head surface points were used for digitisation. Following digitisation, the subject took a seat underneath the MEG device (Neuromag-122; Elekta Neuromag Oy, Helsinki, Finland) which is located in a shielded room (IMEDCO, Hägendorf, Switzerland). Before the start of the first measurement, the subject was instructed to not pay attention to the acoustic stimuli and to instead concentrate on a silent movie (with subtitles) that was projected onto a screen in the wall of the MEG room; the movie was chosen by the subject and served for keeping their vigilance stable. In the final preparatory step, the door of the MEG room was closed and the position of the head and the four coils within the sensor helmet was analysed. Finally, the MEG measurement and the acoustic stimulation were started. During the measurement, the 512 melodies were presented at 70–75 dB SPL. Each melody had a duration of 3600 ms and was separated from the next melody by an inter-stimulus interval of a random length between 1200 and 1250 ms. Neuromagnetic data were acquired using a sampling rate of 1000 Hz and a low-pass filter of 330 Hz. The duration of each MEG run was 41 min 10 sec. After the end of the first MEG run, the subject was offered a break if needed. The second MEG run was carried out in the same way as the first MEG run.

The technical equipment used for the acoustic stimulation comprised a 24-bit sound card (RME ADI 8DS AD/DA interface), an attenuator (Tucker-Davis Technologies PA-5), a headphone buffer (Tucker Davis Technologies HB-7) and Etymotic Research (ER2) earphones that were equipped with 90 cm plastic tubes and foam earpieces. The acoustic delay caused by the plastic tubes was 3.22 ms.

2.1.5 Data Analysis

Pre-Processing of Neuromagnetic Data

All neuromagnetic data were analysed using the programming software MATLAB R2022a (The MathWorks, Inc., Natick, USA) and the source analysis software BESA 5.2. (BESA GmbH, Gräfelfing, Germany). Prior to the actual data analysis, the neuromagnetic data were pre-processed in the following steps: first, the data were inspected manually and bad channels as well as technical and movement artefacts were excluded from data analysis. In a next step, the trigger numbers that stand for the melodies presented during the measurement, were assigned to 21 defined experimental conditions (table 29, table 30). Then, an automatic data scan was performed during which remaining artefacts were detected and excluded from further data analysis; the applied rejection thresholds were 8000 fT/cm (amplitude) and 800 fT/cm/ms (gradient). In the final step of pre-processing, the remaining data were averaged for each of the 21 experimental conditions. The averaged data were finally used for data analysis which included the creation of dipole models and the analysis of the resulting source waveforms and source locations.

Creation of Current Dipole Models

For each subject and MEG run, five different dipole models were created for investigating the neuromagnetic responses to melodies with contour/register/family/sound location change. All dipole models were created using four current dipoles that were inserted into a spherical head model. Two dipoles each were inserted into the left hemisphere and the right hemisphere of the head model. Within each hemisphere, one dipole served for the analysis of the neuromagnetic N1 responses, while the other dipole served for the analysis of the neuromagnetic P2 responses. First, the two N1 dipoles were inserted into the head model and the location of the N1 dipoles was fitted to the neuromagnetic activity at a specific N1 response (or at specific N1 response*s*); then, the N1 dipoles were deactivated, the two P2 dipoles were placed into the head model and the location of the P2 dipoles was fitted to the neuromagnetic activity at a specific P2 response (or at specific P2 response*s*). All dipoles were fitted using a 1 Hz zero-phase high-pass filter. Finally, the two N1 dipoles were activated again and the model was saved. In cases in which one dipole was located outside the temporal lobe, the affected dipole was set symmetric to the corresponding dipole in the other hemisphere. In cases in which the symmetry constraint did not work and one or several dipoles were located outside the temporal lobe or the brain, the dipole model was discarded and not used for further analysis. For melodies with register/family/sound location change, the dipole locations were fitted to the neuromagnetic activity at the change-N1 ($N1_{cng}$) and change-P2 ($P2_{cng}$) responses (i.e. the N1 and P2 responses to the

fourth tone of the melodies). By contrast, for melodies with contour change, the dipole locations were fitted to the neuromagnetic activity at the $N1_{cng}$ and $P2_{cng}$ responses, at the $N1_{post}$ and $P2_{post}$ responses (i.e. the N1 and P2 responses to the fifth tone of the melodies), and at the N1 and P2 responses to the sixth tone of the melodies. This was done because the contour change - which already occurred at the fourth tone of the melodies - could not be detected with certainty until hearing the fifth and the sixth tones as well. By contrast, the register change, family change and location change could already be detected at the same tone where they occurred, i.e. the fourth tone of the melodies.

As indicated by the names of the dipole models, the dipole model CONT was created based on the neuromagnetic responses to all melodies with contour change, the dipole model REG was based on the neuromagnetic responses to all melodies with register change, the dipole model FAM was developed based on the neuromagnetic responses to all melodies with family change, and the dipole model LOC was based on the neuromagnetic responses to all melodies with sound location change. The fifth dipole model ATTACK was created based on the neuromagnetic responses to all melodies, irrespective of the change type. Upon their completion, the four dipole models CONT, REG, FAM and LOC were used as templates for the creation of subsidiary dipole models. Specifically, each template dipole model was applied to create four subsidiary dipole models that had the same dipole locations as the template model but their source waveforms were based on the neuromagnetic responses to a smaller subgroup of melodies with the same change type. For instance, the dipole model CONT served as template for the creation of the subsidiary dipole models $CONT_{OO}$, $CONT_{BB}$, $CONT_{CC}$, and $CONT_{KK}$. All four subsidiary models had the same dipole locations as the CONT model but their source waveforms were based on the neuromagnetic responses to different melody subgroups: the model $CONT_{OO}$ was based on the neuromagnetic responses to melodies with contour change that were played by an oboe, while the model $CONT_{BB}$ was based on the neuromagnetic responses to melodies with contour change that were played by a bassoon, etc. (fig. 7). In total, there were four template dipole models and 16 subsidiary dipole models. As the source locations did not differ between the subsidiary models and the respective template dipole model, the source waveforms were the only result of the subsidiary dipole models that was used for further analysis. By contrast, all results of the template dipole models CONT, REG, FAM and LOC, i.e. the source waveforms and the source locations, were included in the remaining steps of data analysis.

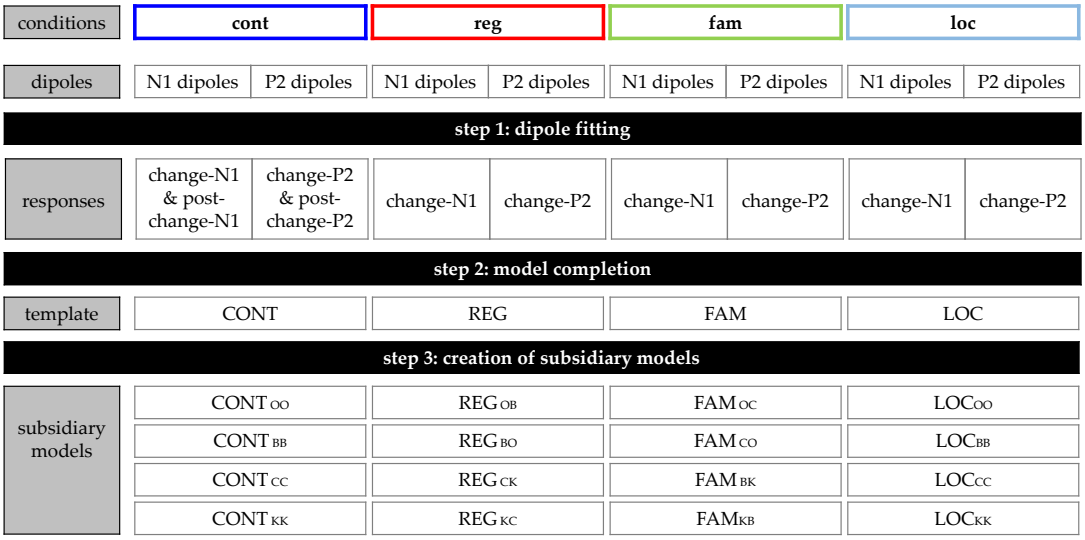


Figure 7: Summary of the dipole modelling process. The pooled experimental conditions contour (cont), register (reg), family (fam) and sound location (loc) and the neuromagnetic responses to the underlying melodies represent the basis for the creation of the dipole models. First, four current dipoles were located into the head model; the location of two dipoles (N1 dipoles) was fitted to (a) specific N1 response(s), whereas the location of the two other dipoles (P2 dipoles) was fitted to (a) specific P2 response(s). In the second step, the temporarily deactivated N1 dipoles were activated again and the completed dipole model was saved. In the third and final step, this dipole model was used as template for the creation of four subsidiary dipole models. The general process for the creation of the dipole models CONT, REG, FAM and LOC also applies to the creation of the ATTACK dipole model (not depicted). However, in contrast to the models CONT, REG, FAM and LOC, the ATTACK model was not used as template for the creation of subsidiary dipole models. Note that all named conditions, template models and subsidiary models refer to both MEG runs A and B. For simplicity, however, the conditions, template models and subsidiary models are only mentioned once using their general names.

Details on the created dipole models are provided in the following sections. Note that the descriptions do not only refer to the dipole models that were built to analyse the results of MEG run A (CONT^{a}, REG^{a}, FAM^{a}, LOC^{a}), but also to the equivalent dipole models that were used to analyse the results of MEG run B (CONT^{b}, REG^{b}, FAM^{b}, LOC^{b}). For simplicity, however, the dipole models are only described once under the general terms CONT, REG, FAM, and LOC.

Dipole model CONT The CONT dipole model was based on the neuromagnetic responses to all melodies with contour change (table 1). Model fitting was performed on the $N1_{cng}$ and $P2_{cng}$ responses, on the $N1_{post}$ and $P2_{post}$ responses, and on the N1 and P2 responses to the sixth tone of the melodies. The CONT model served as template for the creation of the four subsidiary dipole models $CONT_{OO}$,$CONT_{BB}$, $CONT_{CC}$ and $CONT_{KK}$ (table 1).

Dipole model REG The REG dipole model was based on the neuromagnetic responses to all melodies with register change (table 1). Model fitting was only performed on the $N1_{cng}$ and $P2_{cng}$ responses. The REG model served as template for the creation of the four subsidiary dipole models REG_{OB}, REG_{BO}, $REG_{CK}$and REG_{KC} (table 1).

Dipole model FAM The FAM dipole model was based on the neuromagnetic responses to all melodies with family change (table 1). Model fitting was performed on the $N1_{cng}$ and $P2_{cng}$ responses. The FAM model served as template for the creation of the four subsidiary dipole models FAM_{OC}, FAM_{CO}, FAM_{BK} and FAM_{KB} (table 1).

Dipole model LOC The LOC model was based on the neuromagnetic responses to all melodies with sound location change (table 1). Model fitting was performed on the $N1_{cng}$ and $P2_{cng}$ responses. The LOC model served as template for the creation of the four subsidiary dipole models LOC_{OO}, LOC_{BB}, LOC_{CC} and LOC_{KK} (table 1).

Dipole model ATTACK The ATTACK model was based on the neuromagnetic responses to all 32 melodies of a measurement, irrespective of the change type (table 1). Model fitting was performed on the onset-N1 ($N1_{on}$) and onset-P2 ($P2_{on}$) responses. There were no subsidiary dipole models based on the ATTACK model (table 1).

Table 1: Dipole models that were applied for the data analysis of the two MEG runs. For each MEG run, the 21 dipole models comprised five CONT models that were based on the neuromagnetic responses to melodies with contour change (cont), five REG models that were created based on the neuromagnetic responses to melodies with register change (reg), five FAM models that were based on the neuromagnetic responses to melodies with family change (fam), five LOC models that were developed based on the neuromagnetic responses to melodies with sound location change (loc), and one ATTACK model that was created based on the neuromagnetic responses to all melodies (all). The table lists the 2 x 21 dipole models as they were used for statistical analysis, i.e. the corresponding dipole models of MEG run A and MEG run B were pooled to CONT, REG, FAM and LOC. The same applies to the respective underlying experimental conditions. Regarding the ATTACK model, the differentiation between MEG run A and MEG run B was maintained to compare the neuromagnetic responses between the two MEG runs.

Model	Condition	Model	Condition
CONT	**cont**	REG	**reg**
CONT_{OO}	cont_OO	REG_{OB}	reg_OB
CONT_{BB}	cont_BB	REG_{BO}	reg_BO
CONT_{CC}	cont_CC	REG_{CK}	reg_CK
CONT_{KK}	cont_KK	REG_{KC}	reg_KC
FAM	**fam**	LOC	**loc**
FAM_{OC}	fam_OC	LOC_{OO}	loc_OO
FAM_{CO}	fam_CO	LOC_{BB}	loc_BB
FAM_{BK}	fam_BK	LOC_{CC}	loc_CC
FAM_{KB}	fam_KB	LOC_{KK}	loc_KK
ATTACK[a]	**all**	ATTACK[b]	**all**

Statistical Analysis of the Neuromagnetic Data

The data that served as basis for the statistical analyses, comprised the source waveforms and source locations obtained from 320 dipole models. For two subjects, no dipole models could be created due to insufficient data quality; therefore, all data (neuromagnetic data and psychoacoustic data) obtained from the two subjects were excluded from analysis. Of the overall 320 dipole models (10 dipole models for each of the remaining 32 subjects), 41 models (12.81 %) could not be used for further data analysis because the residual variance of the data sets was too high and the resulting dipole models were unstable. Twelve further dipole models (3.75 %) initially were unstable but could be stabilized by applying a 1–30 Hz filter; thus, the 12 dipole models could be maintained for data analysis.

The statistical analysis was performed on the source waveforms and on the source locations. As a non-parametric statistical method that is well-suited for the analysis of physiological data (e.g. Henderson, 2005), bootstrapping was the method of choice for the statistical analysis of the neuromagnetic data. In contrast to other statistical methods like ANOVA (Analysis of Variance), bootstrapping can be applied for statistical data analysis even if dipole modelling was not successful for all experimental conditions in all subjects (e.g. Henderson, 2005). Basically, the bootstrap technique relies on the repeated random sampling of the obtained data and the subsequent data averaging for every sample (DiCiccio and Efron, 1996). The averaged data is then used to build up a data distribution which is finally applied for the calculation of probabilities and confidence intervals (DiCiccio and Efron, 1996). In this study, bootstrap analyses were carried out using 2000 resamples and 95 % BCa (bias-corrected and accelerated) bootstrap intervals. All bootstrap analyses were used for identifying (significant) differences in the N1 and P2 responses to distinct melodies and/or in distinct subject sub-samples (table 31, table 32). Depending on the dipole model, the bootstrap analysis was performed on the $N1_{on}$ and $P2_{on}$ responses (ATTACK dipole model), on the $N1_{cng}$ and $P2_{cng}$ responses (REG, FAM and LOC dipole models), or on the $N1_{post}$ and $P2_{post}$ responses (CONT dipole models).

2.2 Psychoacoustic Experiment

2.2.1 Participants

Seventeen healthy subjects (8 female, 9 male, 31.9 ± 10.9 years, 2 left-handed) participated in the psychoacoustic experiment. Thirteen of the 17 subjects also participated in the MEG experiment.

2.2.2 Stimuli

The stimuli that were used in the psychoacoustic experiment were made up from the same tones as the stimuli that were applied in the MEG experiment. However, in contrast to the stimuli of the MEG experiment, the stimuli of the psychoacoustic experiment were tone pairs consisting of two tones that either differed in pitch, register, family, pitch & register, pitch & family, register & family, or pitch & register & family. Each tone pair consisted of two 600 ms tones that were separated from each other by a 100 ms inter-stimulus interval. The attack time and the sound location were the same for all tones and tone pairs.

2.2.3 Paradigm

The psychoacoustic experiment comprised a short training (ca. 3 min) and a main test (ca. 15 min). Both the training and the test were made up of tone pairs with change in pitch/register/family/pitch & register/pitch & family/register & family/pitch & register & family: the training round consisted of 14 tone pairs (two tone pairs each with the same change type), whereas the test consisted of 168 tone pairs (24 tone pairs each with the same change type) (fig. 8). In the training, the 14 tone pairs were only presented once and in a fixed order. In the test, 168 tone pairs were presented; the 168 tone pairs comprised all 132 possible tone combinations of the applied 12 tones plus 36 repeated tone pairs. Among the repeated tone pairs, there were 12 tone pairs with register change, 12 tone pairs with family change, and 12 tone pairs with register & family change. The reason for why the these tone pairs were presented twice over the course of the measurement, was that there were 24 tone pairs each with change in pitch/pitch & register/pitch & family/pitch & register & family (table 33), but only 12 tone pairs each with change in register/family/register & family (table 34); thus, the 36 tone pairs were presented twice in order to compensate for the lack of data and to obtain the same amount of data for each tone pair group. All 168 tone pairs were presented in a random order that varied from subject to subject.

2.2.4 Measurement

As the rationale behind the psychoacoustic experiment was to investigate the perception of timbre using the sounds' brightness, the subjects were instructed to listen carefully to the tone pairs and to judge for each pair whether the first or the second tone sounded brighter. To ensure that the subjects did not confound brightness with pitch height, the subjects were informed about the difference between pitch height and brightness. The general procedure of the psychoacoustic training and the test was as follows: after the presentation of a tone pair, the subject was asked whether the first or the second tone sounded brighter; the subjects answered by pressing a key and the next tone pair was presented. Both the training and the test were carried out using MATLAB R2022a (The MathWorks, Inc., Natick, USA). Similar to the MEG experiment, the acoustic stimuli were delivered using a 24-bit sound card (RME ADI 8DS AD/DA interface), an attenuator (Tucker-Davis Technologies PA-5) and a headphone buffer (Tucker Davis Technologies HB-7), and were presented via Etymotic Research (ER2) earphones that were equipped with 90 cm plastic tubes and foam earpieces.

2.2.5 Data Analysis

The subjects' brightness judgements were evaluated using a Bradley-Terry analysis (Bradley and Terry, 1952) to obtain a ranking scale for the perceived brightness of the 12 different instrumental tones that were used for the pair comparisons. For each subject, a ranking scale was obtained whose values represented the perceived brightness of a tone in relation to the other 11 tones. Finally, bootstrap analyses were conducted to find out to what extent the ranking values differed **(1)** between the three tones of each instrument (i.e. differences in pitch), **(2)** between the oboe and bassoon, as well as between the clarinet and bass clarinet (i.e. differences in register), **(3)** between the oboe and clarinet, as well as between the bassoon and bass clarinet (i.e. differences in instrument family), **(4)** between the oboe and bass clarinet, as well as between the clarinet and bassoon (i.e. differences in register and instrument family).

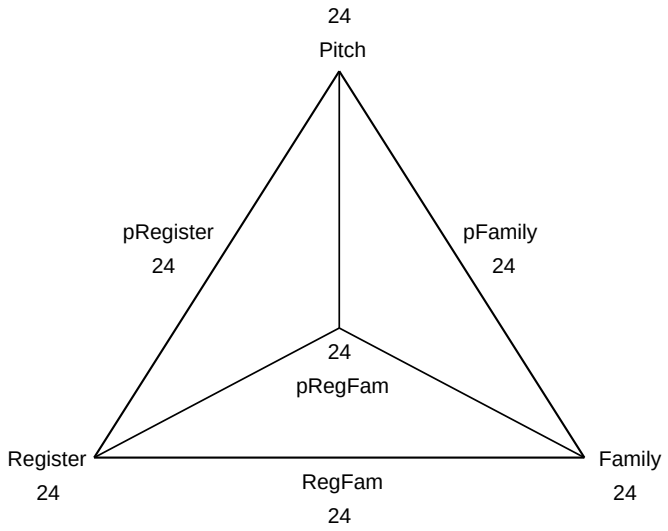


Figure 8: Basic structure of the psychoacoustic test paradigm. The corners of the triangle represent the three groups of tone pairs with only one change type, i.e. tone pairs with pitch change (*Pitch*), tone pairs with register change (*Register*), and tone pairs with family change (*Family*). All other groups of tone pairs are represented by the outer lines and the center of the triangle: the outer lines represent the three groups of tone pairs with two change types, i.e. tone pairs with pitch & register change (*pRegister*), tone pairs with pitch & family change (*pFamily*), and tone pairs with register & family change (*RegFam*). The center of the triangle, where the three inner lines meet, represents the only group of tone pairs with all three change types, i.e. pitch & register & family change (*pRegFam*).

2.3 Sound Features and Combined Analyses

All 12 sound that were used for creating the acoustic stimuli for the MEG experiment and the psychoacoustic experiment (oboe/clarinet: D4, F♯4, F4; bassoon/bass clarinet: D2, F♯2, A2), were analysed regarding their physical characteristics. For each sound, a stabilised auditory image (SAI) was generated which served for the extraction and graphical representation of the sound features periodicity and hissiness. The creation of the SAIs was carried out in MATLAB R2022b, applying a basilar membrane model (transmission line filterbank *physiological route*, neural activity patterns (NAP) model *meddis hair cell*, middle ear filtering (standard variables as described in Patterson et al. (1992) and Patterson et al. (1995)), a sampling frequency of 48,000 Hz, and a frame rate of 10 ms. Finally, the SAIs were completed using a frequency range of 50–15,000 Hz and 500 channels that represent the neural activity within small basilar membrane regions. The length of the SAIs was 40 ms, of which the first 5 ms represent a baseline interval, and the remaining 35 ms show the channel activity along the basilar membrane. Based on the SAIs, each sound's periodicity and hissiness were determined: the periodicity was extracted using the distance between two waveform peaks, and the hissiness was determined by calculating the angle between the zero line and a line through the waveform minima. Finally, a factor analysis was performed between the neuromagnetic data, psychoacoustic data, SAI data and further physical sound properties which were extracted and provided by Christoph Reuter (University of Vienna, Vienna). In total, Christoph Reuter extracted 47 sound properties using the Python-based programme PADMEA (Perceptual Audio Dimension Modelling and Extraction Application; developed by Czedik-Eysenberg (2021) from the University of Vienna). However, the introduction of the 47 sound properties into this study would have gone far beyond the scope of this work; therefore, only the most important and most relevant sound properties (attack time and spectral centroid) were considered in the factor analysis. The factor analysis was conducted in SAS ® OnDemand for Academics (SAS Institute Inc., Cary, North Carolina, USA) and comprised a principal component analysis and a subsequent data rotation using the VARIMAX method.

2.4 Further Psychoacoustic Tests

2.4.1 Audiometric Screening

Prior to participating in the MEG measurements, the subjects were asked to participate in an audiometric screening to rule out hearing disorders. The subjects' task was to listen to tones with different frequencies (125 Hz, 250 Hz, 500 Hz, 750 Hz, 1000 Hz, 1500 Hz, 2000 Hz, 3000 Hz, 4000 Hz, 6000 Hz, 8000 Hz, 12,000 Hz) that were presented at increasing volume and to press a button as soon as the tone could be detected. In case a subject's perception threshold surpassed the maximum acceptable threshold of 30 dB, the respective subject was excluded from subsequent measurements. The acoustic stimuli of the audiometric screening were delivered via an audiometer (OtoCure, Klinik Audiometer KA 350) and presented to the subject via Sennheiser HDA200 headphones.

2.4.2 Assessment of Musical Aptitude

The participants' musical aptitude was assessed by applying the Advanced Measures of Music Audiation (AMMA; in the following named AMMA) test battery which was developed by Gordon (1989). This test battery comprises 30 melody pairs and each melody pair consists of two melodies that may either be identical or that may differ in tone sequence or in rhythm. Accordingly, the subjects' task was to listen to the melody pairs and to judge for each pair whether the two melodies were identical or differed in tone sequence or rhythm. First, the subjects were asked to participate in a short training round to become familiar with the task; then, the test was conducted. After hearing the second melody of the respective melody pair, the subjects were instructed to provide one of three possible answers (melodies are identical, melodies differ in tone sequence, melodies differ in rhythm). At the end of the test, a score of up to 80 points was obtained which provides information on the subject's individual musical aptitude; the higher the musicality score, the higher may be the subject's musical aptitude (Gordon, 1989, pg. 33–34). The melodies were delivered via an amplifier (Lavry Engineering, DA10) and presented via headphones (Sennheiser, HDA200) at moderate volume (ca. 65 dB).

2.4.3 Pitch Perception Test

A pitch perception test (in the following named PITCHT) designed by Schneider et al. (2005) was applied to determine which part of the sound spectrum is used by the subjects to identify the direction of pitch shifts. The test consists of 168 pairs of complex tones that vary in frequency, order and number of harmonics (Schneider et al., 2005). Subjects were instructed to listen to the tone pairs and to judge for

each tone pair whether the second tone was lower or higher than the first tone; the subjects' answers were delivered via key press. At the end of the test, a score between -1 and 1 was obtained which reveals a preference for a pitch perception method: a negative score (< 0) means that the respective subject primarily uses the fundamental or f_0 frequency to detect pitch changes, a positive score (> 0) signifies that the respective subject primarily uses the rest of the spectrum including the overtones. Accordingly, subjects with a negative score are referred to as f_0 *listeners*, while subjects with a positive score are referred to as *overtone listeners*. Subjects with a neutral score (around 0) do not show a preference for any of the two pitch perception methods; they are referred to as *neutral listeners*.
The pitch perception test was performed using MATLAB 7.1. (The MathWorks, Inc., Natick, USA); the melodies were delivered via an amplifier (Lavry Engineering, DA10) and presented via headphones (Sennheiser HDA200) at moderate volume (ca. 65 dB).

2.4.4 Questionnaire on Musical Experience

All participants were asked to fill in a questionnaire on their practical musical experience, their musical activity (no musical activity/music as a hobby/music student/professional musician), their musical training (age at training onset, training duration, training intensity in childhood (< 10 years), youth (10–14 years), adolescence (15–18 years) and in the last six months).

2.4.5 Data Analysis

Using MATLAB, correlation tests were performed to investigate the relation between **(1)** the subjects' age at the date of measurement and the AMMA scores; **(2)** the subjects' PITCHT scores and AMMA scores; **(3)** the subjects' age at training onset and the AMMA scores; **(4)** the subjects' training duration and the AMMA scores; **(5)** the subjects' musical activity and the AMMA scores, and between **(6)** the subjects' age at training onset and the training duration. Additionally, two-sided t-tests were carried out to ensure that the AMMA scores and PITCHT scores did not differ significantly between male and female subjects.

3 Results

3.1 Sample Description

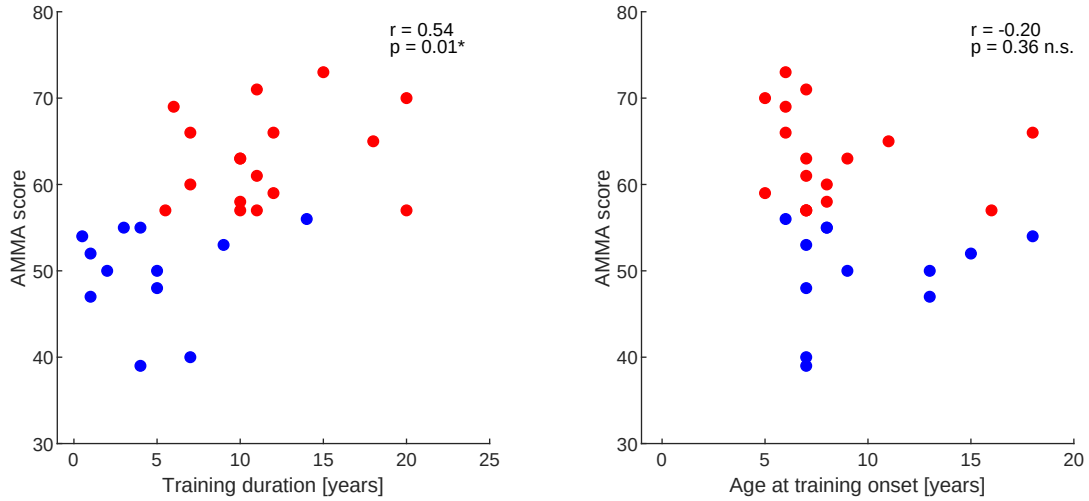
3.1.1 NEURO Subject Group

Musical Aptitude

In the sample of subjects who participated in the MEG experiment and whose data were used for analysis (32 of 34 subjects), the mean AMMA score was 57.3 ± 8.2. The median AMMA score was 57 and was used to split the sample into a *low AMMA* group and a *high AMMA* group; the low AMMA group comprised the subjects with an AMMA score < 57, while the high AMMA group comprised the subjects with an AMMA score ≥ 57. Results of the correlation tests revealed a significant correlation between the training duration and the AMMA score, i.e. the longer the training duration, the higher the AMMA score (fig. 9a). Furthermore, there was a significant correlation between the musical activity and the AMMA score, i.e. the AMMA score rose with increasing musical activity. In contrast to previous studies and the assumption that subjects who started musical training early would also achieve a higher AMMA score, no significant correlation was found between the age at training onset and the AMMA score (fig. 9b). All results of the mentioned correlation tests, as well as the results of correlation tests between the age at the date of measurement and the AMMA score, and between the PITCHT score and the AMMA score, are listed in table 2. The *t*-test between the AMMA scores of male participants and the AMMA scores of female participants did not show significant differences (table 3).

Musical Experience

On average, subjects started musical training at an age of 8.9 ± 3.8 years and took music lessons over a period of 8.7 ± 5.4 years (table 6). There were no significant correlations between the age at the date of measurement and the age at training onset, between the age at the date of measurement and the training duration (table 2), and between the age at training onset and the training duration (fig. 10, table 2). The sample of 32 subjects included 11 musically inactive persons, 18 amateur musicians, and 3 former or actual music students (table 6).



(a) Training duration x AMMA **(b)** Training onset x AMMA

Figure 9: Correlations between the training duration and the AMMA score (**a**) and between the age at training onset and the AMMA score (**b**). Both graphs represent the data obtained from the NEURO subject group. Red dots represent the data of high AMMA listeners, i.e. subjects with an AMMA score at or above the median score (57); blue dots represent the data of low AMMA listeners, i.e. subjects with an AMMA score below the median score. The r- and p- values are provided for the respective correlations.

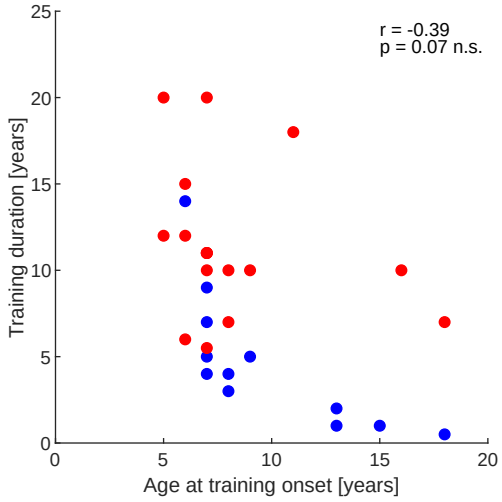


Figure 10: Correlation between the age at training onset and the training duration for the NEURO subject group. Red dots represent the data of high AMMA listeners, i.e. subjects with an AMMA score at or above the median score (57); blue dots represent the data of low AMMA listeners, i.e. subjects with an AMMA score below the median score.

Pitch Perception Type

Of the 32 subjects, 18 obtained a negative PITCHT score and were classified as *fundamental* PITCHT listeners, while 14 subjects obtained a positive PITCHT score and were classified as *overtone* PITCHT listeners. In the group of fundamental PITCHT listeners, the mean score was -0.38 ± 0.23; in the group of overtone PITCHT listeners, the average score was 0.22 ± 0.17. There was no significant correlation between the PITCHT score and the AMMA score (table 2). A significant difference was found between the PITCHT scores of male and female participants (table 3): the PITCHT scores of male subjects were significantly higher than the PITCHT scores of female subjects.

Table 2: Results of the correlation tests between the subjects' age at the date of measurement (*age*), aspects of musical experience – i.e. the subjects' age at training onset (*training onset*), duration of musical training (*training duration*) and musical activity – , the subjects' AMMA scores and PITCHT scores. The data used for the correlation tests were obtained from the NEURO subject group.

Correlation test			
Variable 1	**Variable 2**	***r***	***p***
age	AMMA score	-0.0964	0.5998 n.s.
age	training onset	0.3243	0.1409 n.s.
age	training duration	-0.1943	0.3862 n.s.
PITCHT score	AMMA score	-0.0269	0.8839 n.s.
training onset	AMMA score	-0.2045	0.3613 n.s.
training onset	training duration	-0.3947	0.0691 n.s.
training duration	AMMA score	0.5355	**0.0102***
musical activity	AMMA score	0.5251	**0.0020****

Table 3: Results of the two-sided *t*-tests that were performed to compare the AMMA scores and PITCHT scores between male and female subjects of the NEURO subject group.

	***t*-test**		
Variable	**Group 1**	**Group 2**	***p***
AMMA score	female subjects	male subjects	0.3132 n.s.
PITCHT score	female subjects	male subjects	**0.0019****

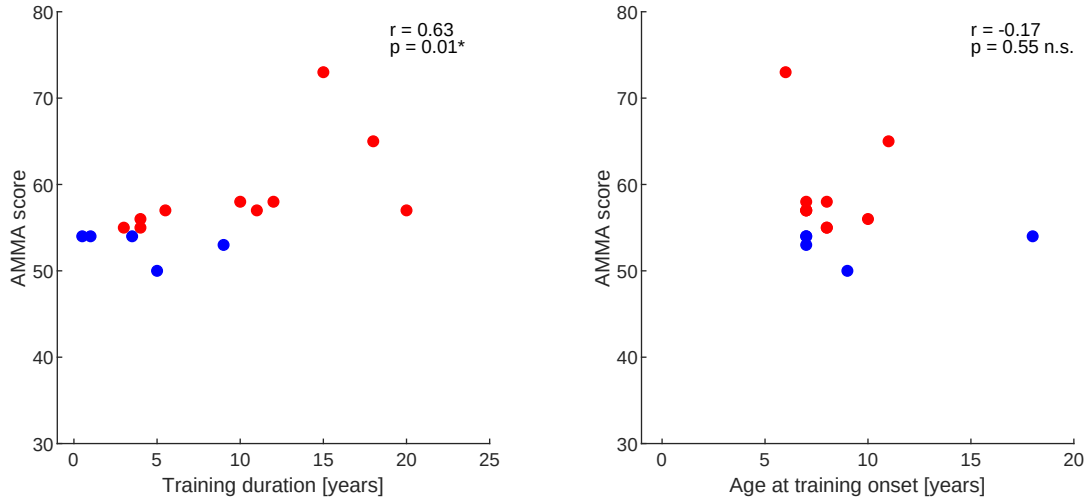
3.1.2 PSYCH Subject Group

Musical Aptitude

In the sample of subjects who participated in the psychoacoustic experiment, the mean AMMA score was 56.0 ± 6.0 and the median AMMA score was 55. As in the NEURO subject group, the median score was used to divide the sample into a *low AMMA* group (AMMA score < 55) and a *high AMMA* group (AMMA score ≥ 55). Results revealed a significant correlation between the training duration and the AMMA score, i.e. the longer the training duration, the higher the AMMA score (fig. 11a). In addition, a significant correlation was found between the musical activity and the AMMA score, i.e. the AMMA score increased with rising musical activity. Against the expectation that a lower training onset age might lead to a significantly higher AMMA score, no significant correlation was found between the age at training onset and the AMMA score (fig. 11b). All results of the mentioned correlation tests, as well as the results of correlation tests between the age at the date of measurement and the AMMA score, and between the PITCHT score and the AMMA score, are listed in table 4. The *t*-test between the AMMA scores of male participants and the AMMA scores of female participants did not show significant differences (table 5).

Musical Experience

On average, subjects started musical training at an age of 8.5 ± 2.9 years and took music lessons for 8.1 ± 6.1 years (table 6). No significant correlations were found between the age at the date of measurement and the age at training onset, between the age at the date of measurement and the training duration (table 4), and between the age at training onset and the training duration (fig. 12, table 4). The PSYCH subject group was composed of 10 musically inactive persons and 7 amateur musicians (table 6). Further data on the PSYCH and NEURO subject groups can be found in table 35.



(a) Training duration x AMMA **(b)** Training onset x AMMA

Figure 11: Correlations between the training duration and the AMMA score (**a**), and between the age at training onset and the AMMA score (**b**). Both graphs represent the data obtained from the PSYCH subject group. Red dots represent the data of high AMMA listeners, i.e. subjects with an AMMA score at or above the median score (55); blue dots represent the data of low AMMA listeners, i.e. subjects with an AMMA score below the median score. The r- and p- values are provided for the respective correlations.

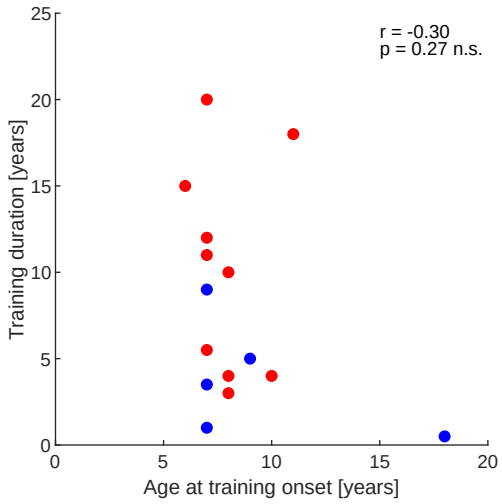


Figure 12: Correlation between the age at training onset and the training duration for the PSYCH subject group. Red dots represent the data of high AMMA listeners, i.e. subjects with an AMMA score at or above the median score (55); blue dots represent the data of low AMMA listeners, i.e. subjects with an AMMA score below the median score.

Pitch Perception Type

Of the 17 subjects who participated in the psychoacoustic experiment, 10 obtained a negative PITCHT score and were classified as *fundamental* PITCHT listeners, whereas 7 subjects obtained a positive PITCHT score and were classified as *overtone* PITCHT listeners (table 6). On average, the PITCHT scores amounted to -0.39 ± 0.25 in the group of fundamental PITCHT listeners and 0.21 ± 0.15 in the group of overtone PITCHT listeners. The PITCHT score was not significantly correlated with the AMMA score (table 4). Furthermore, there was no significant difference between the PITCHT scores of male participants and the PITCHT scores of female participants (table 5).

Table 4: Results of the correlation tests between the subjects' age at the date of measurement (*age*), aspects of musical experience – i.e. the subjects' age at training onset (*training onset*), duration of musical training (*training duration*) and musical activity – , the subjects' AMMA scores and PITCHT scores. The data used for the correlation tests were obtained from the NEURO subject group. The data used for the correlation tests were obtained from the PSYCH subject group.

Correlation test			
Variable 1	**Variable 2**	***r***	***p***
age	AMMA score	-0.3195	0.2112 n.s.
age	training onset	-0.0070	0.9802 n.s.
age	training duration	0.0256	0.9280 n.s.
PITCHT score	AMMA score	0.0045	0.9863 n.s.
training onset	AMMA score	-0.1695	0.5458 n.s.
training onset	training duration	-0.3047	0.2695 n.s.
training duration	AMMA score	0.6263	**0.0125***
musical activity	AMMA score	0.5133	**0.0351***

Table 5: Results of the two-sided *t*-tests that were performed to compare the AMMA scores and PITCHT scores between male and female subjects of the PSYCH subject group.

	***t*-test**		
Variable	**Group 1**	**Group 2**	***p***
AMMA score	female subjects	male subjects	0.9385 n.s.
PITCHT score	female subjects	male subjects	0.2165 n.s.

Table 6: Basic information on the NEURO and PSYCH subject groups.

		NEURO	**PSYCH**
Participants		32 (of 34)	17
female		17	8
male		15	9
Handedness			
left		2	2
right		30	15
Age	[years]	28.1 ± 9.2	31.9 ± 10.9
AMMA score		57.3 ± 8.2	56.0 ± 6.0
median		57	55
$<$ median		14	7
$\geq$ median		18	10
PITCHT score		-0.12 ± 0.36	-0.14 ± 0.37
median		-0.08	-0.08
< 0		18	10
> 0		14	7
Training onset	[age in years]	8.9 ± 3.8	8.5 ± 2.9
Training duration	[time in years]	8.7 ± 5.4	8.1 ± 6.1
Musical activity			
none		11	10
amateur		18	7
student		3	-
professional		-	-

3.2 Neuromagnetic Responses

In the following sections, the neuromagnetic responses to the different melodies will be presented. The neuromagnetic data comprise the source waveforms which represent the neuromagnetic activity per time, and the neural source locations which represent the site from which the detected activity originated.

3.2.1 Source Waveforms

All source waveforms exhibited the same general structure: every single tone of each melody produced an auditory evoked field (AEF) which comprised the three transient responses P1, N1 and P2 and a sustained field. The AEF elicited by the first tone is the pitch-onset response (POR), whereas the AEF evoked at the end of the melody is the pitch-offset response. This study concentrated on the N1 and P2 responses of the first AEF ($N1_{on}$, $P2_{on}$) and fourth AEF ($N1_{cng}$, $P2_{cng}$) (fig. 13). As the MEG data show, the largest transient responses were evoked by the first tone and the fourth tone of the melody. In particular, the P2 responses to the fourth tone were significantly larger than the responses to the second, third, fifth and six tones of the melody (table 36). An overview of the source waveforms elicited by the different melodies is provided in fig. 13.
Figure 14 shows a close-up view of the AEF complex that was elicited by the contour/register/family/sound location change at the fourth tone of the melody. Especially large $N1_{cng}$ and $P2_{cng}$ and responses were evoked by register changes; this effect was more prominent in the $P2_{cng}$ than in the $N1_{cng}$ response. Sound location changes also led to larger $N1_{cng}$ responses, but the $P2_{cng}$ responses remained comparatively small. The smallest $N1_{cng}$ and $P2_{cng}$ responses were elicited by contour changes. Regarding the $N1_{cng}$ response, latency was slightly longer for melodies with family change and melodies with sound location change. With respect to the $P2_{cng}$ response, latency was only longer for melodies with family change. The results of the statistical analysis will be outlined in the following sections.

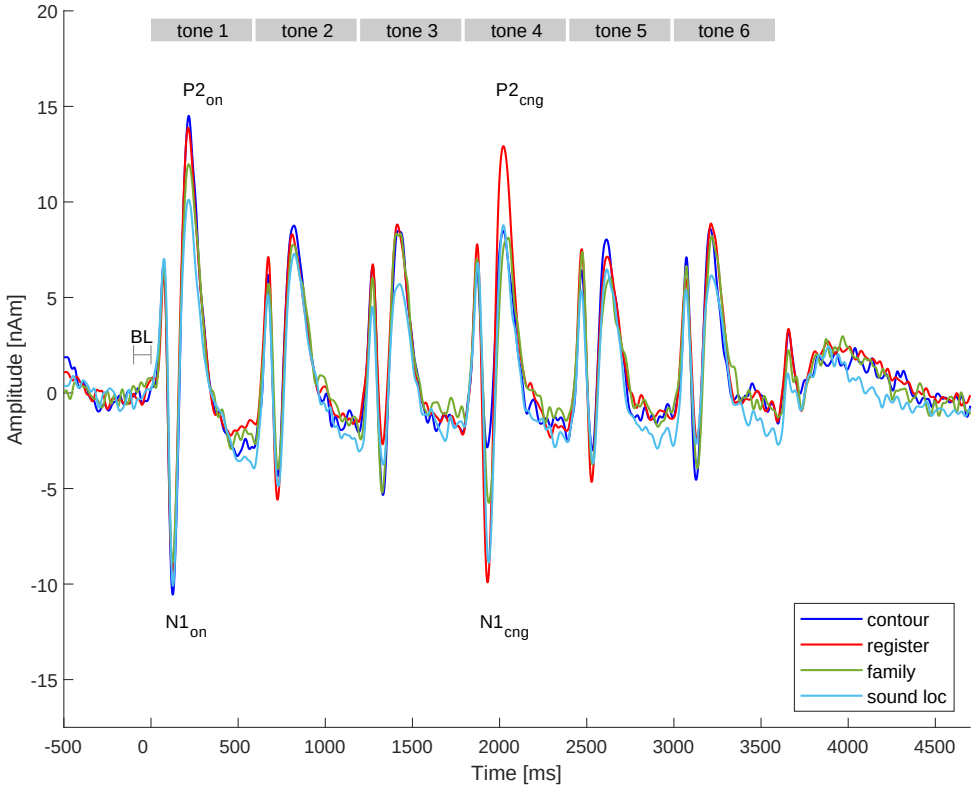


Figure 13: Overview of the source waveforms in response to melodies with contour change (*contour*), register change (*register*), family change (*family*) or sound location (*sound loc*). Source waveforms are displayed as neuromagnetic activity in [nAm] per time in [ms]. The six 600 ms tones of the melody are represented by the grey blocks above the source waveforms. BL = baseline.

Contour In order to find out whether melodies with contour change elicited stronger and/or earlier N1 and P2 responses than melodies without contour change, the amplitudes and latencies of the $N1_{post}$ and $P2_{post}$ responses were statistically analysed. The rationale behind this was that – although the contour change already occurred at the fourth tone of the melody – , the change was more likely to be perceived at the fifth tone because the contour change resembled a simple pitch change which also occurred at the other tone transitions of the melody. Note that the paradigm only included melodies with contour/register/family/sound location change and there was no melody without any change; thus, *melodies without contour change* comprise all melodies with register/family/sound location change and exclude melodies with contour change. As shown in fig. 15 **A**, melodies without contour change evoked larger $N1_{post}$ responses than melodies with contour

change; however, the difference was not significant (table 7). Regarding the latency, melodies without contour change elicited significantly earlier $N1_{post}$ responses than melodies with contour change (table 7). The opposite was observed for the $P2_{post}$ response: the amplitude was significantly larger for melodies with contour change than for melodies without contour change, but the latency did not differ significantly between melodies with and melodies without contour change (fig. 15 **A**, table 7).

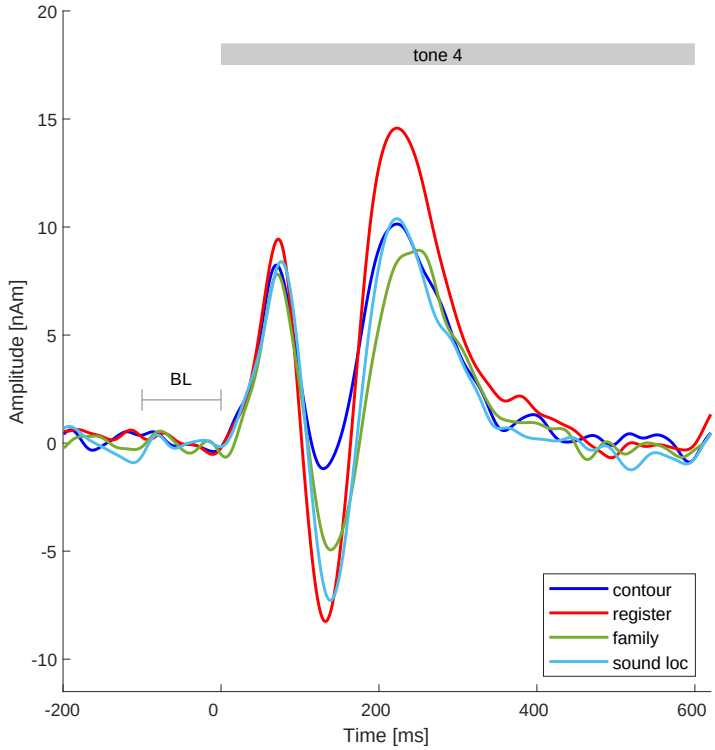


Figure 14: Close-up view of the source waveforms in response to melodies with contour change (*contour*), register change (*register*), family change (*family*) or sound location (*sound loc*). The graph only shows the $N1_{cng}$ and $P2_{cng}$ responses to the fourth tone of the melody. Start, duration and end of the fourth tone are indicated by the grey block above the source waveforms. The graph displays the source waveforms as neuromagnetic activity in [nAm] per time in [ms]. BL = baseline.

Register The register main effect was statistically analysed on the basis of the $N1_{cng}$ and $P2_{cng}$ responses evoked by melodies with register change and melodies without register change. The rationale behind this was that register changes could already be perceived at the tone where they occurred (i.e. the fourth tone of

the melody). Note that *melodies without register change* comprise all melodies with contour/family/sound location change and exclude melodies with register change. The results in fig. 15 **B** show that melodies with register change elicited significantly stronger and earlier $N1_{cng}$ and $P2_{cng}$ responses than melodies without register change (table 7). Regarding the $N1_{cng}$ response, latency was significantly shorter for melodies with register change than for melodies without register change. Concerning the $P2_{cng}$ response, there was no significant latency difference between melodies with and melodies without register change (table 7).

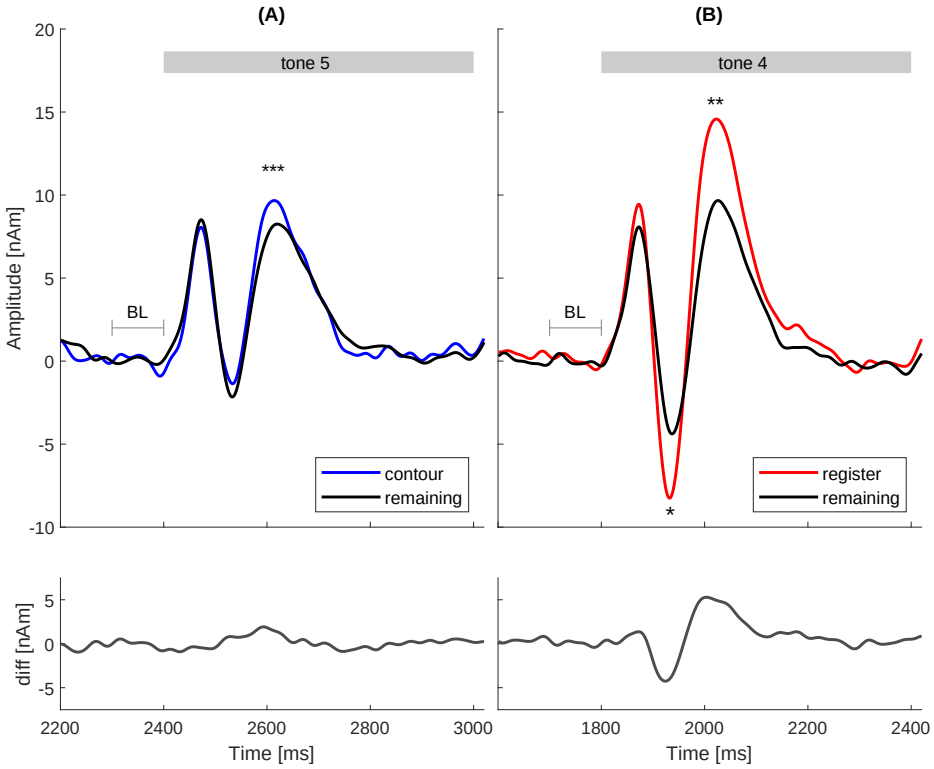


Figure 15: Close-up views of source waveforms (**A**) in response to melodies with contour change (*contour*) and melodies without contour change (*remaining*), and (**B**) in response to melodies with register change (*register*) and melodies without register change (*remaining*). (**A**) shows the $N1_{post}$ and $P2_{post}$ responses to the fifth tone of the melody, (**B**) displays the $N1_{cng}$ and $P2_{cng}$ responses to the fourth tone of the melody. The start, duration and end of the tones are indicated by the grey block above the source waveforms. Asterisks indicate significant amplitude differences. The upper graphs display the source waveforms as neuromagnetic activity in [nAm] per time in [ms]; the lower graphs show the difference of the two source waveforms as neuromagnetic activity in [nAm] per time in [ms]. BL = baseline.

Table 7: Results of the bootstrap analyses on the $N1_{post}$ and $P2_{post}$ responses evoked by melodies with/without contour change (*cont vs rem*), and on the $N1_{cng}$ and $P2_{cng}$ responses evoked by melodies with/without register change (*reg vs rem*).

Comparison	Response	Effect size	*p*
cont vs rem	$N1_{post}$ latency	0.4981	**0.0310***
	$N1_{post}$ amplitude	-0.0414	0.4510 n.s.
	$P2_{post}$ latency	-0.2337	0.1825 n.s.
	$P2_{post}$ amplitude	0.7873	**< 0.001*****
reg vs rem	$N1_{cng}$ latency	-0.5143	**0.0245***
	$N1_{cng}$ amplitude	-0.4679	**0.0325***
	$P2_{cng}$ latency	-0.0460	0.4605 n.s.
	$P2_{cng}$ amplitude	0.7542	**0.0015****

Family Similar to the register main effect, the family main effect was investigated based on the $N1_{cng}$ and $P2_{cng}$ responses because family changes could already be perceived at the fourth tone of the melody. The statistical analysis was performed on the latencies and amplitudes of the $N1_{cng}$ and $P2_{cng}$ responses which were evoked by melodies with family change and melodies without family change. Note that *melodies without family change* include all melodies with contour/register/sound location change and exclude melodies with family change. Regarding the amplitudes of the $N1_{cng}$ and $P2_{cng}$ responses, no significant differences were found between melodies with family change and melodies without family change (table 8). With respect to the latency, results differed between the $N1_{cng}$ response and the $P2_{cng}$ response: the latency of the $N1_{cng}$ response did not differ significantly between melodies with and melodies without family change, but the latency of the $P2_{cng}$ response was significantly shorter for melodies without family change than for melodies with family change (fig. 16 **A**, table 8).

Sound location As changes in the sound location could already be perceived at the fourth tone of the melody, the statistical analysis was performed on the latencies and amplitudes of the $N1_{cng}$ and $P2_{cng}$ responses that were elicited by melodies with sound location change and melodies without sound location change. Note that *melodies without sound location change* comprise all melodies with contour/register/family change and exclude melodies with sound location.
Results of the statistical analysis show that the amplitude of the $N1_{cng}$ response did not differ significantly between melodies with sound location change and melodies without sound location change. However, melodies without sound location change elicited significantly earlier $N1_{cng}$ responses than melodies with sound location

change (fig. 16 **B**, table 8). Melodies without sound location change also elicited significantly stronger $P2_{cng}$ responses than melodies with sound location change. The latency of the $P2_{cng}$ did not differ significantly between melodies with and melodies without sound location change (table 8).

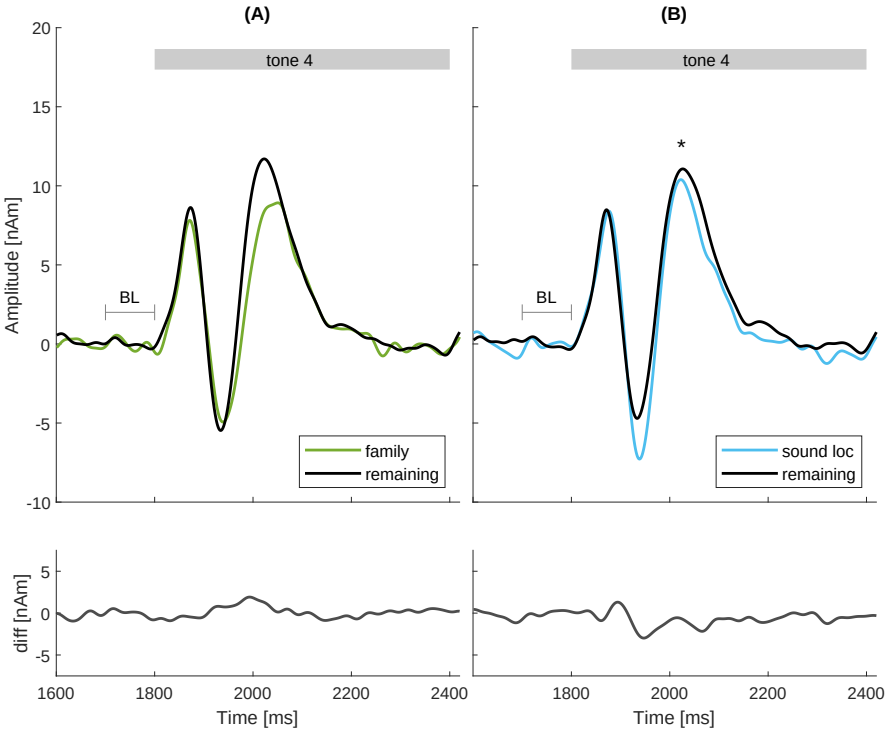


Figure 16: Close-up views of the source waveforms (**A**) in response to melodies with family change (*family*) and melodies without family change (*remaining*), and (**B**) in response to melodies with sound location change (*sound loc*) and melodies without sound location change (*remaining*). Both graphs show the $N1_{cng}$ and $P2_{cng}$ responses to the fourth tone of the melody. The start, duration and end of the tone are indicated by the grey block above the source waveforms. Asterisks indicate significant amplitude differences. The upper graphs display the source waveforms as neuromagnetic activity in [nAm] per time in [ms]; the lower graphs show the difference of the two source waveforms as neuromagnetic activity in [nAm] per time in [ms]. BL = baseline.

Table 8: Results of the bootstrap analyses on the $N1_{cng}$ and $P2_{cng}$ responses evoked by melodies with/without family change (*fam vs rem*), and on the $N1_{cng}$ and $P2_{cng}$ responses evoked by melodies with/without sound location change (*loc vs rem*).

Comparison	Response	Effect size	*p*
fam vs rem	$N1_{cng}$ latency	0.2892	0.1495 n.s.
	$N1_{cng}$ amplitude	0.1149	0.3190 n.s.
	$P2_{cng}$ latency	0.4951	**0.0250***
	$P2_{cng}$ amplitude	-0.0234	0.4610 n.s.
loc vs rem	$N1_{cng}$ latency	0.5270	**0.0330***
	$N1_{cng}$ amplitude	-0.3207	0.1005 n.s.
	$P2_{cng}$ latency	-0.1954	0.2835 n.s.
	$P2_{cng}$ amplitude	-0.4963	**0.0225***

Attack time To investigate the influence of attack time on the neuromagnetic responses, statistical analysis was also performed on the amplitudes and latencies of the $N1_{on}$ and $P2_{on}$ responses to melodies with natural attack times and melodies with unified attack times. The results are displayed in fig. 17 **A** and reveal that melodies with unified attack times elicited significantly earlier and stronger $N1_{on}$ responses, as well as significantly stronger $P2_{on}$ responses than melodies with natural attack times (table 9). The $P2_{on}$ latency was also shorter for melodies with unified attack times than for melodies with natural attack times; however, this effect was not significant (table 9).

Hemisphere Finally, the $N1_{cng}$ and $P2_{cng}$ responses were compared between the left and right cerebral hemispheres, i.e. between the left and right dipoles. To this, the amplitudes and latencies of the $N1_{cng}$ and $P2_{cng}$ responses to all melodies (irrespective of change type and attack time) were examined.
As the results displayed in fig. 17 **B** show, the amplitudes of the $N1_{cng}$ and $P2_{cng}$ responses did not differ significantly between the left hemisphere and right hemisphere (table 9). In contrast to this, the latency of the $N1_{cng}$ response was significantly shorter in the right hemisphere than in the left hemisphere. The latency of the $P2_{cng}$ response did not differ significantly between the hemispheres (table 9).

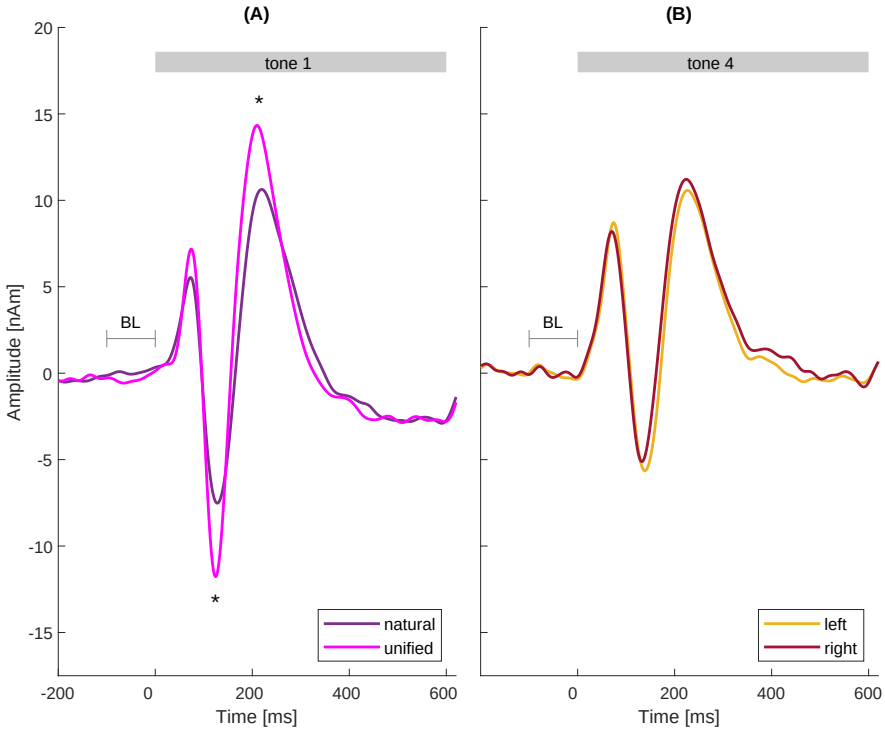


Figure 17: Close-up views of source waveforms. (**A**) shows the source waveforms in response to melodies with natural attack times (*natural*) and melodies with unified attack times (*unified*); (**B**) displays the source waveforms obtained from the left cerebral hemisphere (*left*) and right cerebral hemisphere (*right*). (**A**) depicts the $N1_{on}$ and $P2_{on}$ responses to the first tone of the melody, (**B**) shows the $N1_{cng}$ and $P2_{cng}$ responses to the fourth tone of the melody. In both graphs, the start, duration and end of the respective tone are indicated by the grey block above the source waveforms. Asterisks indicate significant amplitude differences. All source waveforms are displayed as neuromagnetic activity in [nAm] per time in [ms]. BL = baseline.

Table 9: Results of the bootstrap analyses on the $N1_{on}$ and $P2_{on}$ responses evoked by melodies with natural/unified attack times (*nat vs uni*), and on the $N1_{cng}$ and $P2_{cng}$ responses obtained from the left/right cerebral hemispheres (*left vs right*).

Comparison	Response	Effect size	*p*
nat vs uni	$N1_{on}$ latency	0.6843	**0.0040****
	$N1_{on}$ amplitude	0.5041	**0.0260***
	$P2_{on}$ latency	0.4236	0.0550 n.s.
	$P2_{on}$ amplitude	-0.4612	**0.0285***
left vs right	$N1_{cng}$ latency	0.5449	**0.0190***
	$N1_{cng}$ amplitude	-0.0232	0.4575 n.s.
	$P2_{cng}$ latency	0.0235	0.5195 n.s.
	$P2_{cng}$ amplitude	0.1108	0.3285 n.s.

AMMA A further analysis was performed to find out how the subjects' musical aptitude influenced the amplitudes and latencies of the $N1_{cng}$ and $P2_{cng}$ responses. To this end, the $N1_{cng}$ and $P2_{cng}$ responses to all melodies (irrespective of change type and attack time) were compared between the low AMMA and high AMMA listeners of the NEURO subject group.
The results (fig. 18 **A**) reveal that the amplitude of the $N1_{cng}$ responses did not differ significantly between low AMMA and high AMMA listeners. In contrast to this, the $P2_{cng}$ responses were significantly larger for high AMMA than for low AMMA listeners. Latencies of the $N1_{cng}$ and $P2_{cng}$ responses did not differ significantly between low AMMA and high AMMA listeners (table 10).

PITCHT Likewise, a bootstrap analysis was conducted to investigate the influence of the pitch perception type on the amplitudes and latencies of the $N1_{cng}$ and $P2_{cng}$ responses. Again, the analysis was based on the $N1_{cng}$ and $P2_{cng}$ responses to all melodies, irrespective of change type and attack time. The $N1_{cng}$ and $P2_{cng}$ responses were compared between the fundamental PITCHT and overtone PITCHT listeners of the NEURO subject group.
The results are displayed in fig. 18 **B**. It was found that fundamental PITCHT listeners exhibited significantly larger $N1_{cng}$ and $P2_{cng}$ responses than overtone PITCHT listeners. By contrast, latencies of the $N1_{cng}$ and $P2_{cng}$ responses did not differ significantly between fundamental PITCHT and overtone PITCHT listeners (table 10).

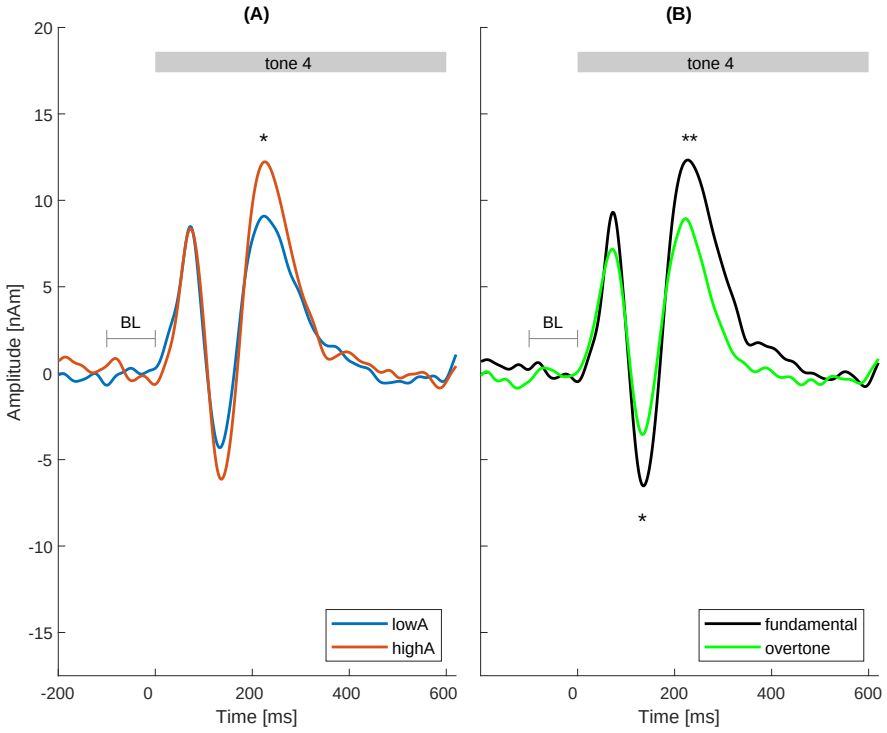


Figure 18: Close-up views of the source waveforms (**A**) obtained from low AMMA listeners (*lowA*) and high AMMA (*highA*) listeners, and (**B**) obtained from fundamental PITCHT listeners (*fundamental*) and overtone PITCHT listeners (*overtone*). The graphs show the $N1_{cng}$ and $P2_{cng}$ responses to the fourth tone of the melody. The start, duration and end of the tone are indicated by the grey block above the source waveforms. Asterisks indicate significant amplitude differences. All source waveforms are displayed as neuromagnetic activity in [nAm] per time in [ms]. BL = baseline.

Interactions with the AMMA score Significant interactions were found (1) between melodies with/without contour change and AMMA scores, (2) between melodies with/without register change and AMMA scores, and (3) between melodies with/without sound location change and AMMA scores. Firstly, the effect that melodies without contour change elicited stronger $N1_{post}$ responses than melodies with contour change, was significantly more pronounced in high AMMA listeners than in low AMMA listeners (fig. 19 **A**, table 11). Secondly, the effect that melodies with register change evoked larger $P2_{cng}$ responses than melodies without register change, was significantly stronger in high AMMA listeners than in low AMMA listeners (fig. 19 **B**, table 11). Finally, the effect that melodies without sound location change produced stronger $P2_{cng}$ responses than melodies with sound

location change, was significantly more pronounced in high AMMA listeners than in low AMMA listeners (fig. 20 **B**, table 11). No significant interactions were observed between melodies with/without family change and the AMMA scores (fig. 20 **A**, table 11).

Table 10: Results of the bootstrap analyses on the $N1_{cng}$ and $P2_{cng}$ responses obtained from low AMMA and high AMMA listeners (*lowA vs highA*), and on the $N1_{cng}$ and $P2_{cng}$ responses obtained from fundamental PITCHT and overtone PITCHT listeners (*fund vs over*).

Comparison	**Response**	**Effect size**	***p***
lowA vs highA	$N1_{cng}$ latency	0.1910	0.3810 n.s.
	$N1_{cng}$ amplitude	0.5203	0.0865 n..s
	$P2_{cng}$ latency	0.2124	0.2625 n.s.
	$P2_{cng}$ amplitude	-0.6977	**0.0190***
fund vs over	$N1_{cng}$ latency	0.1456	0.3920 n.s.
	$N1_{cng}$ amplitude	-0.6604	**0.0220***
	$P2_{cng}$ latency	0.0070	0.4615 n.s.
	$P2_{cng}$ amplitude	0.9157	**0.0035****

Table 11: Results of the bootstrap analyses on interactions between the $N1_{on}$ and $P2_{on}$ responses to melodies with/without contour change and the AMMA score (*cont x AMMA*), between the $N1_{cng}$ and $P2_{cng}$ responses to melodies with/without register change and the AMMA score (*reg x AMMA*), between the $N1_{cng}$ and $P2_{cng}$ responses to melodies with/without family change and the AMMA score (*fam x AMMA*), and between the $N1_{cng}$ and $P2_{cng}$ responses to melodies with/without sound location change and the AMMA score (*loc x AMMA*).

Comparison	**Response**	**Effect size**	***p***
cont x AMMA	$N1_{post}$ amplitude	-0.8121	**0.0175***
	$P2_{post}$ amplitude	-0.2439	0.2595 n.s.
reg x AMMA	$N1_{cng}$ amplitude	0.1617	0.3265 n.s.
	$P2_{cng}$ amplitude	-0.7588	**0.0170***
fam x AMMA	$N1_{cng}$ amplitude	0.0858	0.3915 n.s.
	$P2_{cng}$ amplitude	0.1444	0.3395 n.s.
loc x AMMA	$N1_{cng}$ amplitude	-0.3518	0.1640 n.s.
	$P2_{cng}$ amplitude	0.6342	**0.0355***

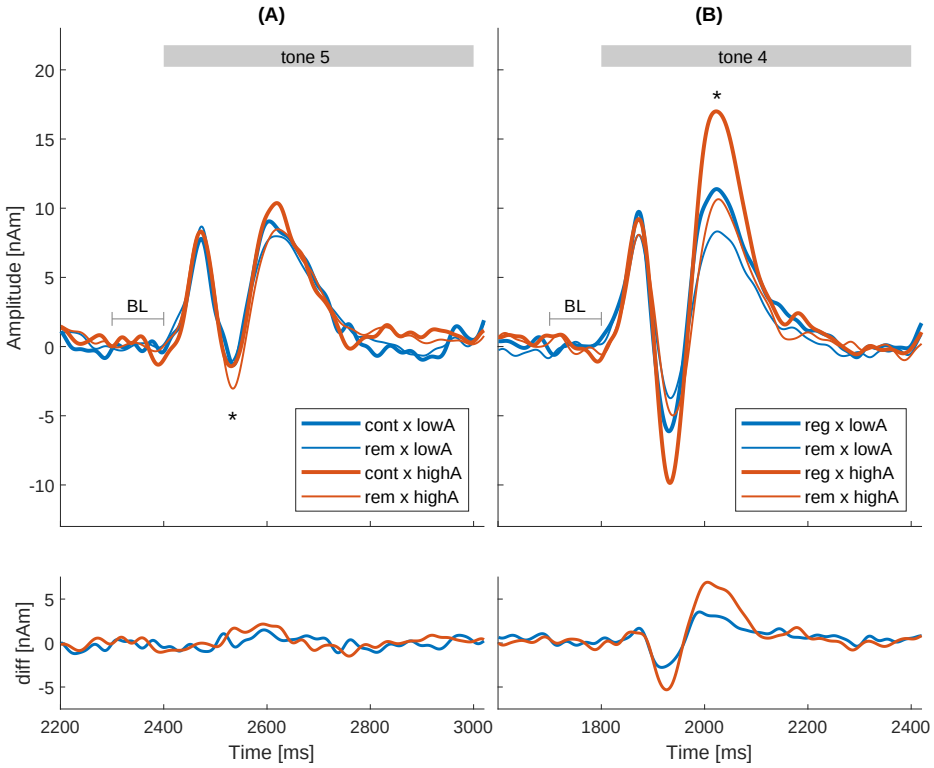


Figure 19: Close-up views of the source waveforms obtained from low AMMA listeners and high AMMA listeners in response to melodies with/without contour or register changes. (**A**) shows the $N1_{post}$ and $P2_{post}$ responses to melodies with contour change (*cont x lowA*, *cont x highA*) and to melodies without contour change (*rem x lowA*, *rem x highA*). (**B**) depicts the $N1_{cng}$ and $P2_{cng}$ responses to melodies with register change (*reg x lowA*, *reg x highA*) and to melodies without register change (*rem x lowA*, *rem x highA*). The start, duration and end of the respective tone are indicated by the grey block above the source waveforms. Asterisks indicate significant amplitude differences. All source waveforms are displayed as neuromagnetic activity in [nAm] per time in [ms]. BL = baseline.

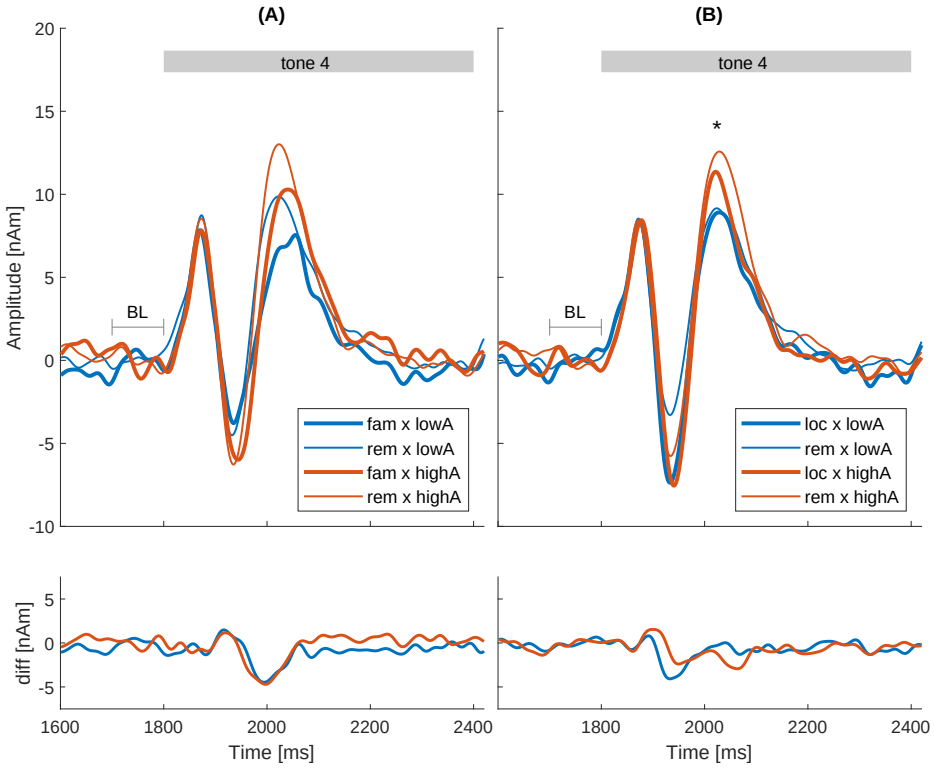


Figure 20: Close-up views of the source waveforms obtained from low AMMA listeners and high AMMA listeners in response to melodies with/without family or sound location changes. (**A**) shows the $N1_{cng}$ and $P2_{cng}$ responses to melodies with family change (*fam x lowA*, *fam x highA*) and to melodies without family change (*rem x lowA*, *rem x highA*). (**B**) depicts the $N1_{cng}$ and $P2_{cng}$ responses to melodies with sound location change (*loc x lowA*, *loc x highA*) and to melodies without sound location change (*rem x lowA*, *rem x highA*). The start, duration and end of the tone are indicated by the grey block above the source waveforms. Asterisks indicate significant amplitude differences. All source waveforms are displayed as neuromagnetic activity in [nAm] per time in [ms]. BL = baseline.

Interaction with the PITCHT score Significant interactions were found (1) between melodies with/without contour change and the PITCHT score, (2) between melodies with/without register change and the PITCHT score, (3) between melodies with/without family change and the PITCHT score, and (4) between melodies with/without sound location change and the PITCHT score. Firstly, the effect that melodies with contour change evoked larger $P2_{post}$ responses than melodies without contour change, was significantly more pronounced in fundamental PITCHT listeners than in overtone PITCHT listeners (fig. 21 **A**, table 12). Secondly, the effect that melodies with register change produced stronger $N1_{cng}$ responses than melodies without register change, was significantly more pronounced in fundamental PITCHT listeners than in overtone PITCHT listeners (fig. 21 **B**, table 12). Thirdly, the effect that melodies without family change elicited more pronounced $P2_{cng}$ responses than melodies with family change, was significantly larger in fundamental PITCHT listeners than in overtone PITCHT listeners (fig. 22 **A**, table 12). Finally, the effect that melodies with sound location change evoked stronger $N1_{cng}$ responses than melodies without sound location change, was more pronounced in fundamental PITCHT listeners than in overtone PITCHT listeners (fig. 22 **B**, table 12).

Table 12: Results of the bootstrap analyses on interactions between the $N1_{on}$ and $P2_{on}$ responses to melodies with/without contour change and the PITCHT score (*cont x PITCHT*), between the $N1_{cng}$ and $P2_{cng}$ responses to melodies with/without register change and the PITCHT score (*reg x PITCHT*), between the $N1_{cng}$ and $P2_{cng}$ responses to melodies with/without family change and the PITCH score (*fam x PITCHT*), and between the $N1_{cng}$ and $P2_{cng}$ responses to melodies with/without sound location change and the PITCHT score (*loc x PITCH*).

Comparison	**Response**	**Effect size**	***p***
cont x PITCHT	$N1_{post}$ amplitude	0.4418	0.1025 n.s.
	$P2_{post}$ amplitude	0.7753	**0.0090****
reg x PITCHT	$N1_{cng}$ amplitude	-0.9489	**0.0020****
	$P2_{cng}$ amplitude	-0.1607	0.3410 n.s.
fam x PITCHT	$N1_{cng}$ amplitude	-0.0677	0.4340 n.s.
	$P2_{cng}$ amplitude	0.6184	**0.0370***
loc x PITCHT	$N1_{cng}$ amplitude	0.6517	**0.0410***
	$P2_{cng}$ amplitude	-0.4496	0.1145 n.s.

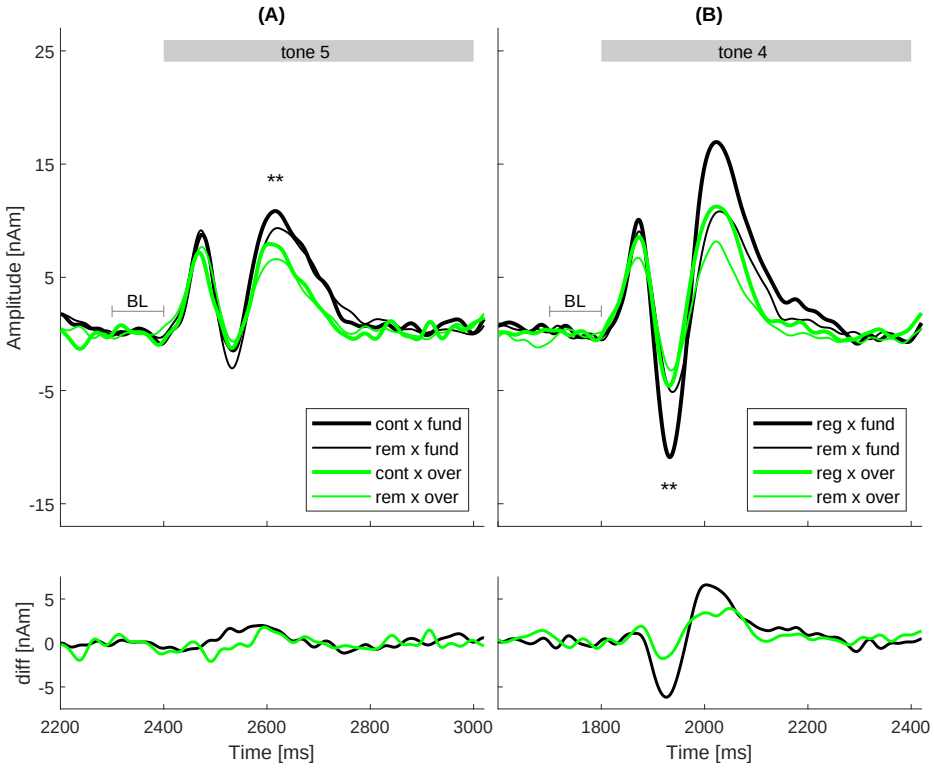


Figure 21: Close-up views of the source waveforms obtained from fundamental PITCHT listeners and overtone PITCHT listeners in response to melodies with/without contour or register changes. (**A**) shows the $N1_{post}$ and $P2_{post}$ responses to melodies with contour change (*cont x fund*, *cont x over*) and to melodies without contour change (*rem x fund*, *rem x over*). (**B**) depicts the $N1_{cng}$ and $P2_{cng}$ responses to melodies with register change (*reg x fund*, *reg x over*) and to melodies without register change (*rem x fund*, *rem x over*). The start, duration and end of the respective tone are indicated by the grey block above the source waveforms. Asterisks indicate significant amplitude differences. All source waveforms are displayed as neuromagnetic activity in [nAm] per time in [ms]. BL = baseline.

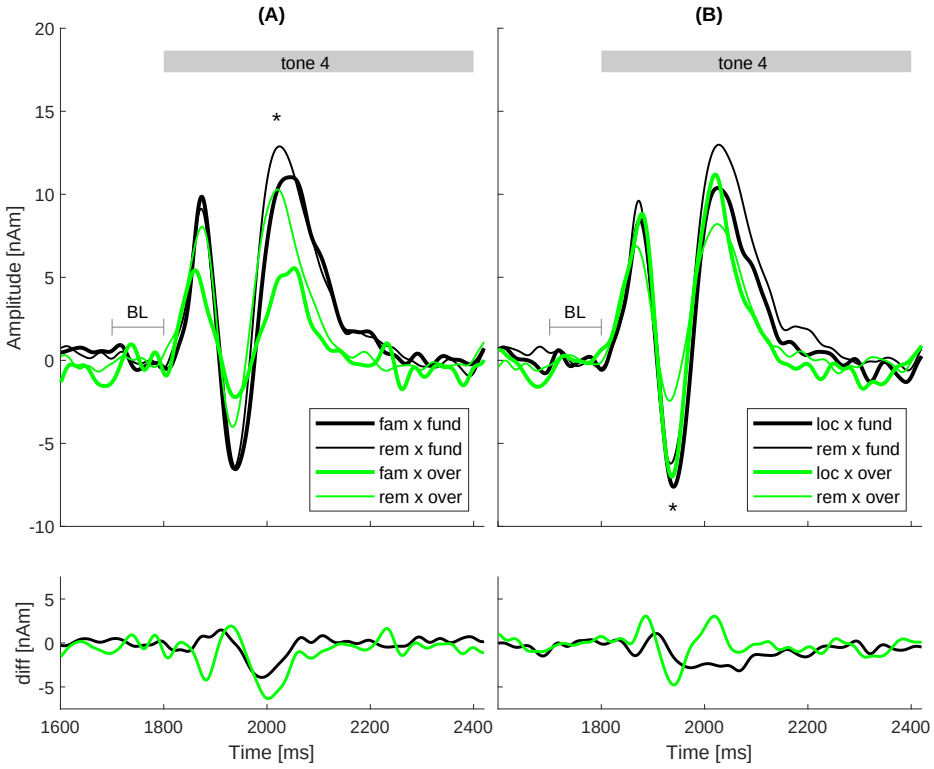


Figure 22: Close-up views of the source waveforms obtained from fundamental PITCHT listeners and overtone PITCHT listeners in response to melodies with/without family or sound location changes. **(A)** shows the $N1_{cng}$ and $P2_{cng}$ responses to melodies with family change (*fam x fund*, *fam x over*) and to melodies without family change (*rem x fund*, *rem x over*). **(B)** depicts the $N1_{cng}$ and $P2_{cng}$ responses to melodies with sound location change (*loc x fund*, *loc x over*) and to melodies without sound location change (*rem x fund*, *rem x over*). The start, duration and end of the tones are indicated by the grey block above the source waveforms. Asterisks indicate significant amplitude differences. All source waveforms are displayed as neuromagnetic activity in [nAm] per time in [ms]. BL = baseline.

3.2.2 Neural Source Locations

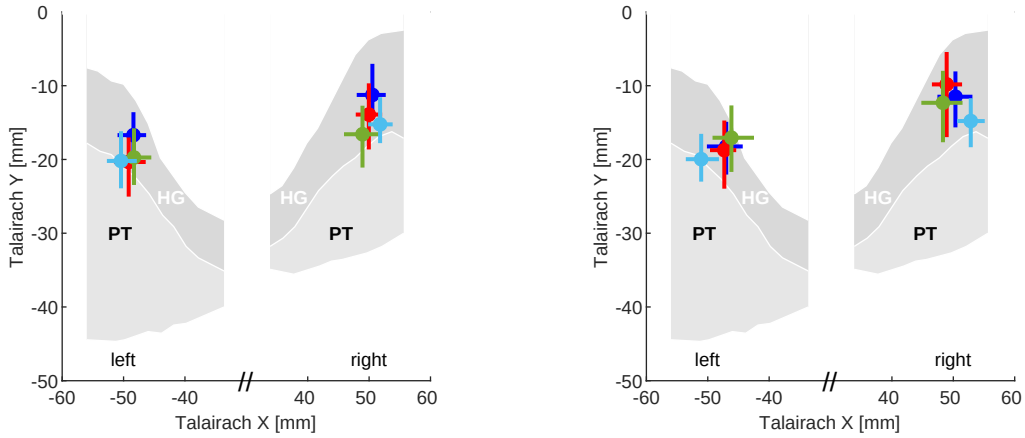
The analysed source locations for the detected $N1_{cng}$, $P2_{cng}$, $N1_{post}$ and $P2_{post}$ responses were displayed on an anatomical map of the bilateral auditory cortex (based on Leonard et al. (1998)). Results of the source analyses and details on the differences between the experimental conditions will be provided in the following sections.

N1 sources Figure 23a depicts the locations of the $N1_{cng}$ and $N1_{post}$ sources. In the left hemisphere, the most posterior neural sources were the $N1_{cng}$ source for melodies with register change and the $N1_{cng}$ source for melodies with sound location change; both neural sources were located near or at the border between the HG and the PT. By contrast, the $N1_{cng}$ source for melodies with family change and the $N1_{post}$ source for melodies with contour change were located more anterior in the HG. The most anterior N1 source was that for melodies with contour change. In the right hemisphere, the $N1_{cng}$ source for melodies with family change was the most posterior; like the $N1_{cng}$ source for melodies with sound location change, it was located in the posterior part of the HG. As in the left hemisphere, the most anterior source was the $N1_{post}$ source for melodies with contour change. The $N1_{cng}$ source for melodies with register change was located between the $N1_{cng}$ source for melodies with sound location change and the $N1_{post}$ source for melodies with contour change. Further important observations were made when comparing the $N1_{cng}$ source locations between low AMMA and high AMMA listeners, as well as between fundamental PITCHT and overtone PITCHT listeners. Regarding the left hemisphere, the $N1_{cng}$ sources were located more posterior in low AMMA listeners and overtone PITCHT listeners, and more anterior in high AMMA listeners and fundamental PITCHT listeners (fig. 24a). The most posterior source was that for the $N1_{cng}$ responses in low AMMA listeners and was located in the PT, close to the border with the HG. In overtone PITCHT listeners, the $N1_{cng}$ source was located at the border between PT and HG. In both the high AMMA listeners and fundamental PITCHT listeners, the $N1_{cng}$ source was located in the HG. The most anterior source was found in fundamental PITCHT listeners. Regarding the right hemisphere, a similar pattern was observed; however, in this case, the most posterior $N1_{cng}$ source was found in overtone PITCHT listeners (near the border between PT and HG) and all $N1_{cng}$ sources were located in the HG. The most anterior source was again that for the $N1_{cng}$ responses in fundamental PITCHT listeners.

The upper section of table 13 shows the results of the bootstrap analyses on the coordinates of the N1 sources. The source coordinates differed significantly between melodies with contour change and melodies without contour change (y-coordinate), between melodies with sound location change and melodies without

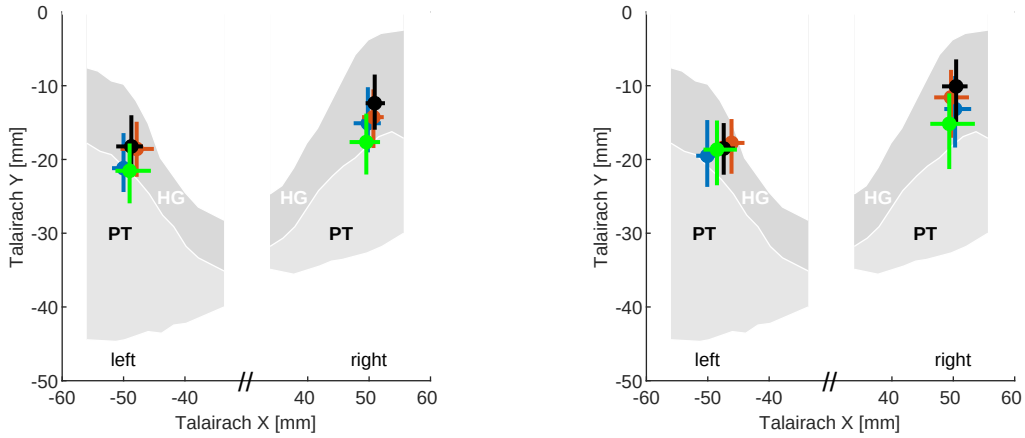
sound location change (z-coordinate), between low AMMA and high AMMA listeners (x-coordinate), and between fundamental PITCHT and overtone PITCHT listeners (y- and z-coordinates).

P2 sources The analysed locations of the $P2_{cng}$ and $P2_{post}$ sources are shown in fig. 23b. In both hemispheres, the most posterior source was the $P2_{cng}$ source for melodies with sound location change. In the left hemisphere, the respective neural source was located at the border between the HG and the PT. The $P2_{cng}$ source for melodies with register change and the $P2_{post}$ source for melodies with contour change were located more anterior in the HG and lay close to each other. Most anterior was the $P2_{cng}$ source for melodies with family change. In the right hemisphere, the sources lay further apart from each other than in the left hemisphere. The most posterior source – the $P2_{cng}$ source for melodies with sound location change – was located near but not at the border between the HG and the PT. Other than in the left hemisphere, the most anterior source in the right hemisphere was the $P2_{cng}$ source for melodies with register change. Both the $P2_{post}$ source for melodies with contour change and the $P2_{cng}$ source for melodies with family change were located posterior to the most anterior source.
Regarding the $P2_{cng}$ source locations in low/high AMMA and fundamental/overtone PITCHT listeners, it was found that the $P2_{cng}$ sources were generally located more anterior than the $N1_{cng}$ sources; this especially applied to the left hemisphere (fig. 24b) where – in contrast to the $N1_{cng}$ sources – no $P2_{cng}$ source was located in the PT. Interestingly, in both hemispheres, all $P2_{cng}$ sources were located in the HG along a line from posterior to anterior. The most posterior $P2_{cng}$ source in the left hemisphere was found in low AMMA listeners, whereas the most anterior source was found in high AMMA listeners. By contrast, in the right hemisphere, the most posterior source was that for the $P2_{cng}$ responses in overtone PITCHT listeners, and the most anterior source was that for the $P2_{cng}$ responses in fundamental PITCHT listeners. Bootstrap results on the coordinates of the P2 sources are given in the lower section of table 13. Results show that the source coordinates differed significantly between melodies with sound location change and melodies without sound location change (z-coordinate), and between low AMMA and high AMMA listeners (x-coordinate).



(a) N1 sources **(b)** P2 sources

Figure 23: Source locations for the N1 and P2 responses to melodies with contour/register/family/sound location change. Anatomical maps (based on Leonard et al. (1998)) show an axial cut through bilateral auditory cortex and cover the Heschl's gyrus (HG) and planum temporale (PT). The x- and y-axes represent the Talairach coordinates X and Y in [mm], respectively. (**a**) depicts the location of the $N1_{post}$ sources for melodies with contour change (blue) and the location of the $N1_{cng}$ sources for melodies with register change (red), melodies with family change (green) and melodies with sound location change (light blue). (**b**) displays the location of the $P2_{post}$ sources for melodies with contour change (blue) and the location of the $P2_{cng}$ sources for melodies with register change (red), melodies with family change (green) and melodies with sound location change (light blue). In both graphs, the circles indicate the average source locations and the lines represent 95% confidence (BCa) intervals.



(a) Change-N1 sources

(b) Change-P2 sources

Figure 24: Source locations of the N1 and P2 responses obtained from low AMMA listeners, high AMMA listeners, fundamental PITCHT listeners and overtone PITCHT listeners. Anatomical maps (based on Leonard et al. (1998)) show an axial cut through bilateral auditory cortex and cover the Heschl's gyrus (HG) and planum temporale (PT). The x- and y-axes represent the Talairach coordinates X and Y in [mm], respectively. The graphs depict the source locations for the $N1_{cng}$ responses (**a**) and $P2_{cng}$ responses (**b**) obtained from low AMMA listeners (blue), high AMMA listeners (red), fundamental PITCHT listeners (black) and overtone PITCHT listeners (green). Both graphs represent average source locations (circles) with 95% confidence (BCa) intervals (lines). The underlying bootstrap analyses were performed on the coordinates of the $N1_{cng}/P2_{cng}$ responses to all melodies.

Table 13: Results of the bootstrap analyses on the x-, y- and z-coordinates of the N1 sources (upper section of the table) and P2 sources (lower section of the table). Results are provided only for the main effects, i.e. contour change vs register/family/sound location changes (*cont vs rem*), register change vs contour/family/sound location changes (*reg vs rem*), family change vs contour/register/sound location changes (*fam vs rem*), sound location change vs contour/register/family changes (*loc vs rem*), natural attack time vs unified attack time (*nat vs uni*), low AMMA listeners vs high AMMA listeners (*lowA vs highA*), and fundamental PITCHT listeners vs overtone PITCHT listeners (*fund vs over*). Regarding the comparison *cont vs rem*, bootstrap analysis was performed on the coordinates of the $N1_{post}$ and $P2_{post}$ sources; concerning the comparison *nat vs uni*, bootstrap analysis was performed on the coordinates of the $N1_{on}$ and $P2_{on}$ sources. For all other comparisons, bootstrap analyses were performed on the coordinates of the $N1_{cng}$ and $P2_{cng}$ sources.

	x		***y***		***z***	
Comparison	**Effect size**	***p***	**Effect size**	***p***	**Effect size**	***p***
cont vs rem	0.1682	0.2565 n.s.	0.4890	**0.0250***	-0.0285	0.4540 n.s.
reg vs rem	-0.1830	0.2285 n.s.	-0.0123	0.4980 n.s.	-0.3119	0.1145 n.s.
fam vs rem	-0.1689	0.2475 n.s.	-0.1844	0.2310 n.s.	-0.1677	0.2535 n.s.
loc vs rem	0.0197	0.4860 n.s.	-0.0405	0.4170 n.s.	0.7382	**0.0030****
nat vs uni	0.2631	0.1445 n.s.	0.1231	0.3110 n.s.	0.2950	0.1285 n.s.
lowA vs highA	-0.5922	**0.0425***	-0.2296	0.2490 n.s.	-0.4147	0.1275 n.s.
fund vs over	0.2589	0.2405 n.s.	0.5874	**0.0420***	0.6460	**0.0335***
cont vs rem	0.1972	0.2330 n.s.	0.0963	0.3545 n.s.	-0.4217	0.0575 n.s.
reg vs rem	-0.1188	0.3170 n.s.	0.0959	0.3435 n.s.	-0.0269	0.4550 n.s.
fam vs rem	0.0205	0.4690 n.s.	0.1583	0.2595 n.s.	-0.0886	0.3710 n.s.
loc vs rem	-0.0519	0.4135 n.s.	-0.1893	0.2245 n.s.	0.7212	**0.0045****
nat vs uni	0.2927	0.1235 n.s.	-0.3323	0.0970 n.s.	-0.3986	0.0620 n.s.
lowA vs highA	-0.6277	**0.0350***	-0.1812	0.3100 n.s.	-0.2855	0.2105 n.s.
fund vs over	0.4539	0.1060 n.s.	0.3482	0.1620 n.s.	0.5434	0.0620 n.s.

3.3 Perceived Brightness of Instrument Tones

Figure 25 shows the results of the Bradley-Terry analysis that was performed on the subjects' brightness judgements for the 12 instrumental tones (oboe/clarinet: D4, F♯4, A4; bassoon/bass clarinet: D2, F♯2, A2). The results reveal that subjects judged the highest-pitched oboe tone A4 and the highest-pitched clarinet tone A4 as brightest and the lowest-pitched bass clarinet tone D2 as least bright. In general, higher pitched tones were judged brighter than lower pitched tones, and the relative brightness values differed significantly between the tones: in each instrument, the relative brightness was highest for the highest-pitched tone A, lower for the middle-pitched tone F♯, and lowest for the lowest-pitched tone D (table 14). Another important finding was that high register tones (i.e. oboe tones and clarinet tones) were generally judged brighter than low register tones (i.e. bassoon tones and bass clarinet tones). In fact, the relative brightness values were significantly higher for high register tones than for low register tones (table 14). The same was observed when comparing the relative brightness of tones that differed in both register and pitch (table 15). Regarding instrument family, similar brightness judgements were obtained for the oboe tones and clarinet tones, and the same effect was observed for the bassoon tones and bass clarinet tones (fig. 25). The relative brightness values did not differ significantly between the oboe family instruments (oboe, bassoon) and the clarinet family instruments (clarinet, bass clarinet) (table 14); however, significant differences were observed when comparing the relative brightness of tones that differed in both family and pitch: in this case, the higher pitched tones were judged significantly brighter than the lower pitched tones (table 15). Finally, significant brightness differences were also observed between tones differing in the two variables register and family or in the three variables pitch, register and family: in both cases, tones from high register instruments (oboe, clarinet) were judged significantly brighter than low register instruments (bassoon, bass clarinet) (table 16).

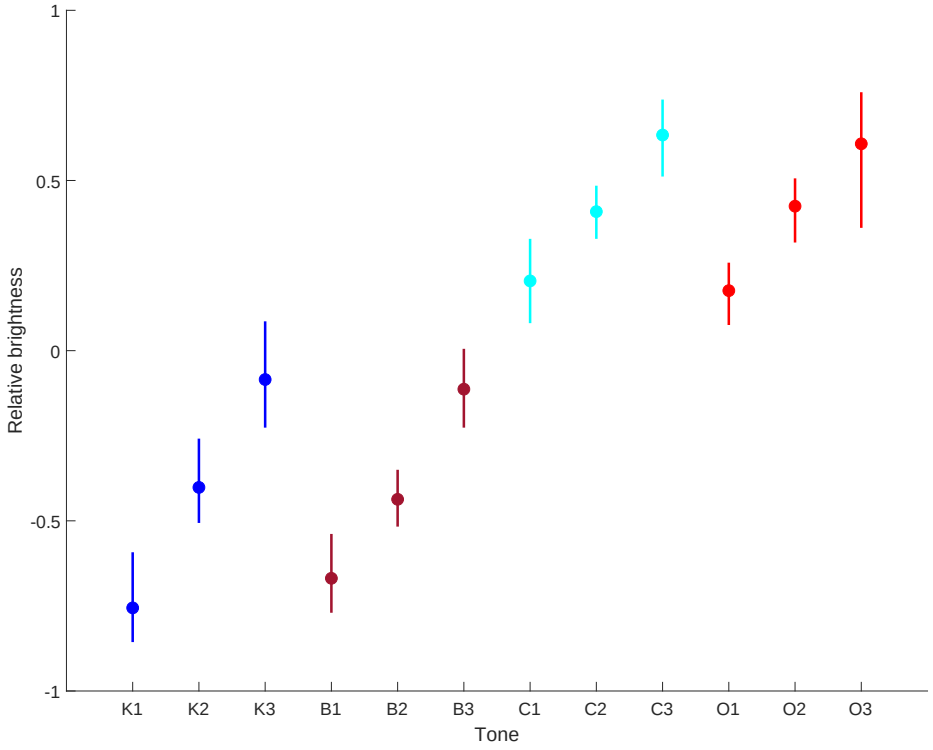


Figure 25: Results of the psychoacoustic experiment, shown as the perceived relative brightness for the 12 presented instrumental tones. For each tone, the relative brightness is provided as an average value between -1 and 1 (circles), with 95 % confidence (Bca) intervals (lines). Dark blue circles and lines represent the relative brightness for the bass clarinet tones D2, F♯2 and A2 (indicated as K1, K2 and K3); dark red circles and lines display the relative brightness for the bassoon tones D2, F♯2 and A2 (given as B1, B2 and B3); light blue circles and lines represent the relative brightness for the clarinet tones D4, F♯4 and A4 (indiciated as C1, C2 and C3), and light red circles and lines display the relative brightness for the oboe tones D4, F♯4 and A4 (given as O1, O2 and O3).

Table 14: Bootstrap results on the relative brightness values for differences in pitch (*Pitch*), differences in register (*Register*) and differences in family (*Family*).

Type of change	**Comparison**	**Effect size**	***p***
Pitch	K1 vs K2	-1.2908	$<$ **0.001*** **
	K1 vs K3	-2.3363	$<$ **0.001*** **
	K2 vs K3	-1.0907	$<$ **0.001*** **
	B1 vs B2	-1.0962	**0.0020** **
	B1 vs B3	-2.2775	$<$ **0.001*** **
	B2 vs B3	-1.5458	$<$ **0.001*** **
	C1 vs C2	-0.9308	**0.0025** **
	C1 vs C3	-1.7246	$<$ **0.001*** **
	C2 vs C3	-1.0774	**0.0015** **
	O1 vs O2	-1.3135	$<$ **0.001*** **
	O1 vs O3	-1.3761	$<$ **0.001*** **
	O2 vs O3	-0.5987	**0.0490***
Register	K1 vs C1	3.6478	$<$ **0.001*** **
	K2 vs C2	3.6535	$<$ **0.001*** **
	K3 vs C3	2.5564	$<$ **0.001*** **
	B1 vs O1	3.8188	$<$ **0.001*** **
	B2 vs O2	4.8084	$<$ **0.001*** **
	B3 vs O3	2.1891	$<$ **0.001*** **
Family	K1 vs B1	0.3265	0.1680 n.s.
	K2 vs B2	-0.1491	0.3345 n.s.
	K3 vs B3	-0.0982	0.3845 n.s.
	C1 vs O1	-0.1123	0.3675 n.s.
	C2 vs O2	0.1001	0.3950 n.s.
	C3 vs O3	-0.0667	0.4420 n.s.

Table 15: Bootstrap results on the relative brightness values for differences in pitch and register (*pRegister*) and differences in pitch and family (*pFamily*).

Type of change	**Comparison**	**Effect size**	***p***
pRegister	K1 vs C2	-5.2457	$<$ **0.001*** **
	K1 vs C3	-5.4859	$<$ **0.001*** **
	K2 vs C3	-4.1466	$<$ **0.001*** **
	B1 vs O2	-5.1484	$<$ **0.001*** **
	B1 vs O3	-3.8929	$<$ **0.001*** **
	B2 vs O3	-3.5306	$<$ **0.001*** **
	C1 vs K2	2.3047	$<$ **0.001*** **
	C1 vs K3	0.9988	**0.0020** **
	C2 vs K3	1.9762	$<$ **0.001*** **
	O1 vs B2	3.4336	$<$ **0.001*** **
	O1 vs B3	1.2955	$<$ **0.001*** **
	O2 vs B3	2.4325	$<$ **0.001*** **
pFamily	K1 vs B2	-1.4136	$<$ **0.001*** **
	K1 vs B3	-2.5215	$<$ **0.001*** **
	K2 vs B3	-1.1214	$<$ **0.001*** **
	B1 vs K2	-1.0502	$<$ **0.001*** **
	B1 vs K3	-2.0590	$<$ **0.001*** **
	B2 vs K3	-1.3848	$<$ **0.001*** **
	C1 vs O2	-0.9303	**0.0055** **
	C1 vs O3	-1.1735	**0.0025** **
	C2 vs O3	-0.6501	**0.0375***
	O1 vs C2	-1.2991	$<$ **0.001*** **
	O1 vs C3	-2.1278	$<$ **0.001*** **
	O2 vs C3	-1.0034	**0.0020** **

Table 16: Bootstrap results on the relative brightness values for differences in register and family (*RegFam*) and differences in pitch, register and family (*pRegFam*).

Type of change	Comparison	Effect size	*p*
RegFam	K1 vs O1	4.1230	$<$ **0.001*****
	K2 vs O2	3.5891	$<$ **0.001*****
	K3 vs O3	1.9305	$<$ **0.001*****
	B1 vs C1	-3.4349	$<$ **0.001*****
	B2 vs C2	-5.0330	$<$ **0.001*****
	B3 vs C3	-3.1475	$<$ **0.001*****
pRegFam	K1 vs O2	-4.9993	$<$ **0.001*****
	K1 vs O3	-3.9376	$<$ **0.001*****
	K2 vs O3	-2.9778	$<$ **0.001*****
	B1 vs C2	-5.1918	$<$ **0.001*****
	B1 vs C3	-5.5046	$<$ **0.001*****
	B2 vs C3	-5.1967	$<$ **0.001*****
	C1 vs B2	2.9037	$<$ **0.001*****
	C1 vs B3	1.2216	$<$ **0.001*****
	C2 vs B3	2.4532	$<$ **0.001*****
	O1 vs K2	2.4891	$<$ **0.001*****
	O1 vs K3	1.0068	**0.0035****
	O2 vs K3	1.9502	$<$ **0.001*****

3.4 Physical Properties of Instrument Tones

In total, four physical sound properties were analysed for each instrumental tone: hissiness and periodicity (which were extracted from the stabilised auditory images, figures 41–52), attack time and spectral centroid (which were extracted by Christoph Reuter (University of Vienna, Vienna) using the software PADMEA). Figure 26 shows that hissiness was lowest for the three bassoon tones and highest for the oboe tone F♯4; periodicity was shortest for the oboe tone/clarinet tone A4 and longest for the bassoon/bass clarinet tone D2. In sum, the hissiness and periodicity values differed more between low register instruments (bassoon, bass clarinet) and high register instruments (clarinet, oboe) than between instruments of the same register.

Similarly, the spectral centroid was lower for the low register instruments than for the high register instruments; the spectral centroid was lowest for the bassoon tone/bass clarinet tone D2 and highest for the clarinet tone D4 (fig. 27). Attack time was shortest for the oboe tone A4 and longest for the bass clarinet tone D2. Interestingly, the attack times of the bass clarinet tones were longer than the attack times of the other instrumental tones (fig. 27).

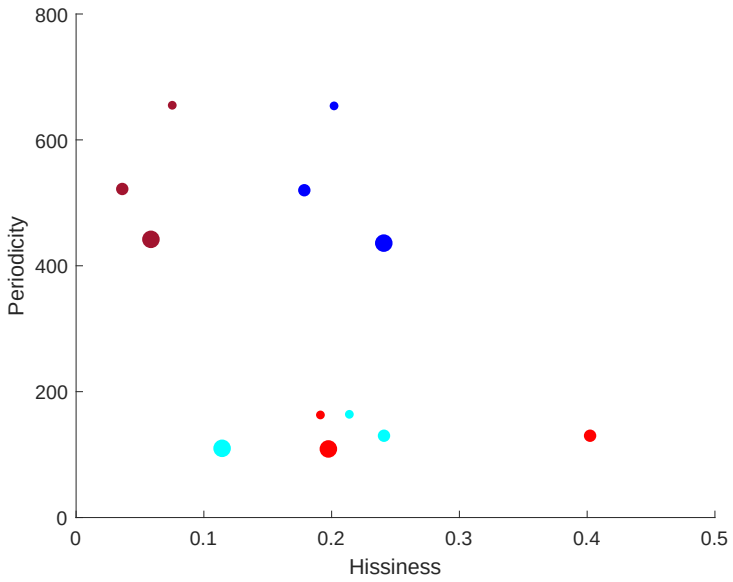


Figure 26: Graphical representation of the hissiness and periodicity value ranges. Hissiness is represented on the x-axis, periodicity is given on the y-axis. Dark blue dots represent the data for the bass clarinet tones D2 (small dot), F♯2 (middle-sized dot) and A2 (big dot); dark red dots depict the data for the bassoon tones D2 (small dot), F♯2 (middle-sized dot) and A2 (big dot); light blue dots represent the data for the clarinet tones D4 (small dot), F♯4 (middle-sized dot) and A4 (big dot), and light red dots depict the data for the oboe tones D4 (small dot), F♯4 (middle-sized dot) and A4 (big dot).

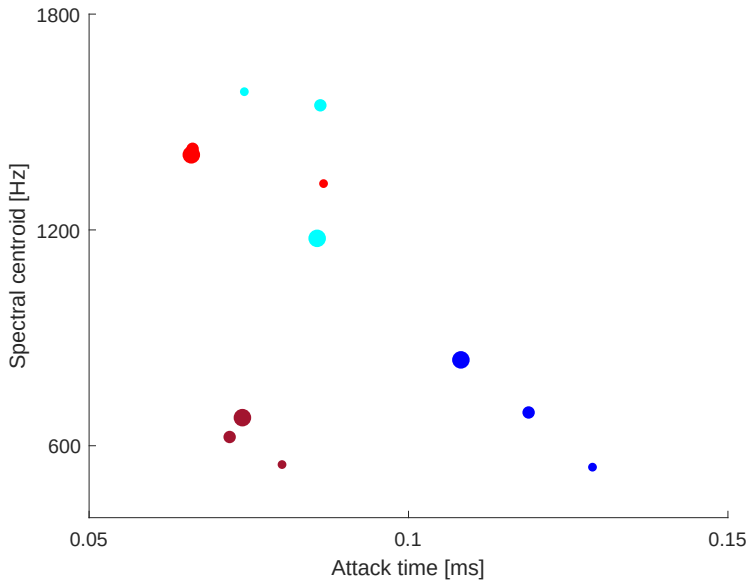


Figure 27: Graphical representation of the attack time and spectral centroid value ranges. The attack time is represented on the x-axis in [ms], the spectral centroid is given on the y-axis in [Hz]. Dark blue dots represent the data for the bass clarinet tones D2 (small dot), F♯2 (middle-sized dot) and A2 (big dot); dark red dots depict the data for the bassoon tones D2 (small dot), F♯2 (middle-sized dot) and A2 (big dot); light blue dots represent the data for the clarinet tones D4 (small dot), F♯4 (middle-sized dot) and A4 (big dot), and light red dots depict the data for the oboe tones D4 (small dot), F♯4 (middle-sized dot) and A4 (big dot).

3.5 Results of Combined Analyses

3.5.1 Neuromagnetic Data and Psychoacoustic Data

Individual tones A major aim of the combined analysis of neuromagnetic data and psychoacoustic data was to find out whether and to what extent the latencies and amplitudes of the N1 and P2 responses to the 12 instrumental sounds were correlated with the perceived brightness of the tones. To this, correlation tests were performed between the subjects' average N1/P2 latencies and amplitudes and their average brightness judgements for the 12 tones (table 17). Results of the correlation tests reveal that the N1 amplitude and the P2 latency were not significantly correlated with the perceived brightness of the tones. However, there was a significant correlation between the N1 latency and relative brightness, as well as between the P2 amplitude and relative brightness. More specifically, brighter tones evoked significantly earlier N1 responses and larger P2 responses than less bright tones (fig. 28 **A, C**). Figure 28 further shows that, interestingly, less bright tones elicited stronger N1 responses than brighter tones (**B**), but brighter tones evoked earlier P2 responses than less bright tones (**D**); however, these effects were not significant (table 17).

Table 17: Results of the correlation tests between the neuromagnetic data obtained from the NEURO subject group and the psychoacoustic data obtained from the PSYCH subject group. The correlation tests were performed between the subjects' average N1 and P2 latencies/amplitudes and their average brightness judgements for each of the 12 instrumental tones.

Variable	r	p
N1 latency	-0.8781	$<$ **0.001*****
N1 amplitude	0.5548	0.061 n.s.
P2 latency	-0.7915	**0.0022****
P2 amplitude	0.7661	**0.0037****

Individual subjects Another aim of the combined analysis of the neuromagnetic data and psychoacoustic data was to investigate whether the N1 and P2 latencies and amplitudes were correlated with the subjects' individual brightness judgements. The rationale behind this was to find out to what extent the neuromagnetic responses and brightness judgements differed between the subjects, and to exclude subject group-related effects on the brightness judgements. Fortunately, no significant correlations were found, neither between the N1 latency/amplitude and relative brightness, nor between the P2 latency/amplitude and relative brightness (table 18). Interestingly, the N1 latencies and brightness judgements of the individual subjects built confined clusters (fig. 29 **A**), whereas the N1 amplitudes, P2 latencies and P2 amplitudes were more widely distributed among the subjects (fig. 29 **B**–**D**).

Table 18: Results of the correlation tests between the neuromagnetic data and psychoacoustic data obtained from subjects who formed part of both the AMMA group *and* the PSYCH group. The correlation tests were performed between the subjects' individual N1 and P2 latencies/amplitudes and their individual brightness judgements, both averaged across the 12 instrumental tones.

Variable	r	p
N1 latency	-0.2406	0.4512 n.s.
N1 amplitude	-0.4657	0.1271 n.s.
P2 latency	-0.1507	0.6401 n.s.
P2 amplitude	0.3491	0.2660 n.s.

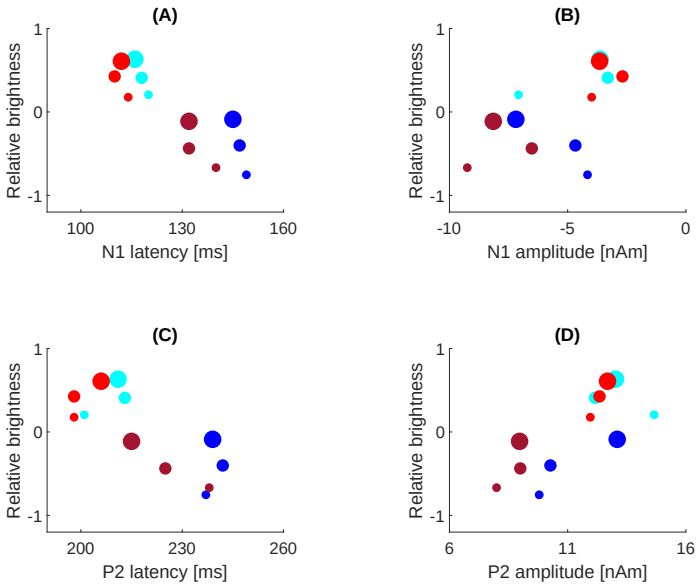


Figure 28: Combined display of the neuromagnetic data obtained from the NEURO subject group and psychoacoustic data obtained from the PSYCH subject group. The neuromagnetic data comprised the average latencies and amplitudes of the subjects' N1 and P2 responses to each of the 12 instrumental tones; the psychoacoustic data contained the subjects' average brightness judgements for the 12 tones. N1 and P2 latencies and amplitudes are represented on the x-axis in [ms] (latency) and [nAm] (amplitude); the relative brightness is represented on the y-axis as a value between -1 and 1. (**A**) and (**B**) display the relative brightness in relation with the N1 latency and N1 amplitude, respectively; (**C**) and (**D**) show the relative brightness in relation with the P2 latency and P2 amplitude, respectively. Note that *larger* N1 responses have *lower* amplitude values. In each subplot, dark blue dots represent the data for the bass clarinet tones D2 (small dot), F♯2 (middle-sized dot) and A2 (big dot); dark red dots depict the data for the bassoon tones D2 (small dot), F♯2 (middle-sized dot) and A2 (big dot); light blue dots represent the data for the clarinet tones D4 (small dot), F♯4 (middle-sized dot) and A4 (big dot), and light red dots depict the data for the oboe tones D4 (small dot), F♯4 (middle-sized dot) and A4 (big dot).

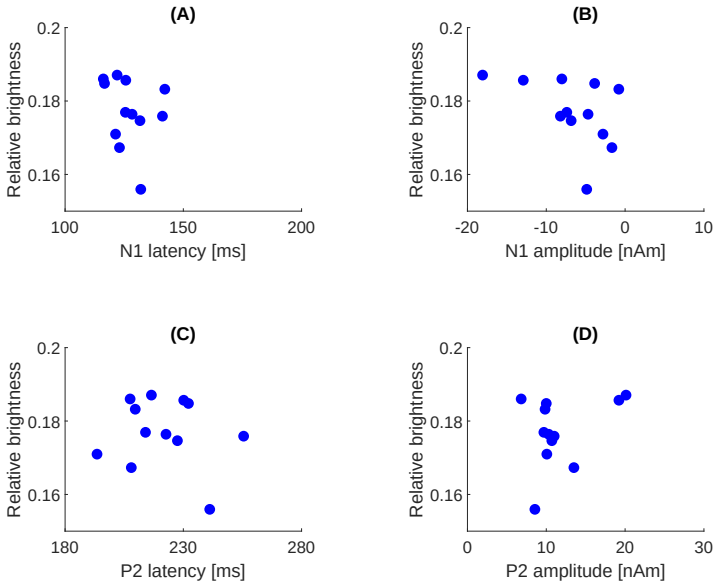


Figure 29: Combined display of the neuromagnetic data and psychoacoustic data, both obtained from subjects who formed part of the NEURO group *and* the PSYCH group. The neuromagnetic data comprised the latencies and amplitudes of the subjects' individual N1 and P2 responses across the 12 instrumental tones; the psychoacoustic data contained the subjects' individual brightness judgements, averaged across the 12 tones. Each dot represents the data obtained from one of the 12 NEURO+PSYCH subjects. N1 and P2 latencies and amplitudes are represented on the x-axis in [ms] (latency;) and [nAm] (amplitude); the relative brightness is shown on the y-axis as a value between -1 and 1. (**A**) and (**B**) display the relative brightness in relation with the N1 latency and N1 amplitude, respectively; (**C**) and (**D**) show the relative brightness in relation with the P2 latency and P2 amplitude, respectively. Note that *larger* N1 responses have *lower* amplitude values.

3.5.2 Neuromagnetic Data, AMMA Scores and PITCHT Scores

AMMA scores In order to exclude further subject group-related effects, correlation tests were performed between the N1 and P2 latencies/amplitudes and the subjects' individual AMMA scores. Results show that the N1 amplitude was slightly correlated with the AMMA score when considering both low AMMA *and* high AMMA listeners (the N1 amplitude significantly increased with rising AMMA score), but not when considering either low AMMA listeners *or* high AMMA listeners (table 19). Regarding the N1 latency, P2 latency and P2 amplitude, no significant correlations were found, neither for all listeners, nor for low AMMA listeners or high AMMA listeners.

Figure 30 reveals that both the low AMMA data and high AMMA data built clusters, although the clusters were denser for the N1 latency and amplitude (**A, B**) than for the P2 latency and amplitude (**C, D**).

Table 19: Results of the correlation tests between the neuromagnetic data and the AMMA score, both obtained from the NEURO subject group. The correlation tests were performed between the subjects' individual AMMA scores and brightness judgements, averaged across the 12 instrumental tones. AMMA = all listeners; low AMMA = only low AMMA listeners; high AMMA = only high AMMA listeners.

Variable	**AMMA**		**low AMMA**		**high AMMA**	
	r	***p***	***r***	***p***	***r***	***p***
N1 latency	-0.2466	0.1737 n.s.	-0.3673	0.1964 n.s.	-0.0014	0.9955 n.s.
N1 amplitude	-0.3567	**0.0450***	-0.5022	0.0673 n.s.	-0.1421	0.5738 n.s.
P2 latency	0.2090	0.2510 n.s	0.2133	0.4640 n.s.	-0.1457	0.5642 n.s.
P2 amplitude	0.2891	0.1085 n.s.	0.3434	0.2293 n.s.	0.2068	0.4104 n.s.

PITCHT scores Finally, correlation tests were performed between the N1 and P2 latencies and amplitudes and the subjects' individual PITCHT scores. Two slightly significant correlations were found: the P2 amplitude significantly increased with decreasing PITCHT score (all listeners), and the P2 latency significantly rose with decreasing PITCHT score (only fundamental PITCHT listeners). No significant correlations were found neither between the N1 latency and the PITCHT score, nor between the N1 amplitude and the PITCHT score. The subjects' individual data resembled each other more regarding the N1 latency (fig. 31, **A**), and less regarding the N1 amplitude, P2 latency and P2 amplitude (fig. 31, **B**–**D**).

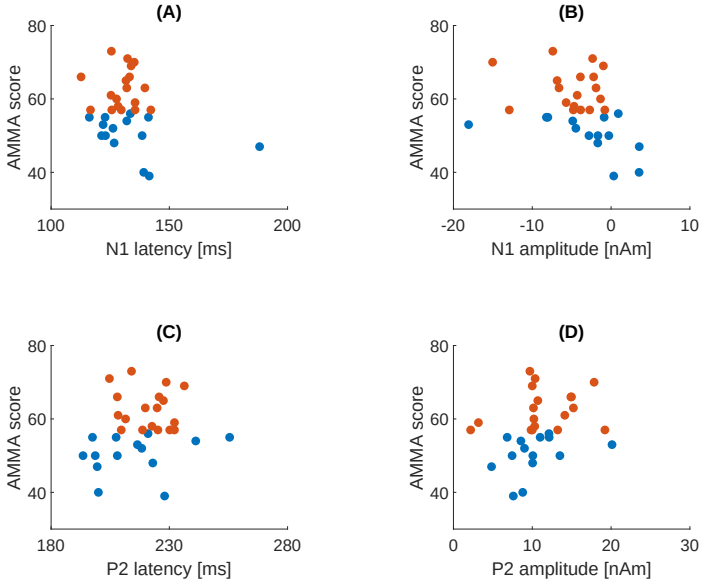


Figure 30: Combined display of the neuromagnetic data and AMMA scores, both obtained from the NEURO subject group. The neuromagnetic data comprised the latencies and amplitudes of the subjects' individual N1 and P2 responses across the 12 instrumental tones and are represented on the x-axis in [ms] (latency) and [nAm] (amplitude). The y-axis shows the subjects' individual AMMA scores as absolute values. Each dot represents the data obtained from one of the 32 NEURO subjects. (**A**) and (**B**) depict the AMMA score in relation with the N1 latency and N1 amplitude, respectively; (**C**) and (**D**) show the AMMA score in relation with the P2 latency and P2 amplitude, respectively. Note that *larger* N1 responses have *lower* amplitude values. In each subplot, blue dots indicate low AMMA subjects and red dots indicate high AMMA subjects.

Table 20: Results of the correlation tests between the neuromagnetic data and the PITCHT score, both obtained from the NEURO subject group. The correlation tests were performed between the subjects' individual PITCHT scores and brightness judgements, averaged across the 12 instrumental tones. PITCHT = all listeners; fund. PITCHT = only fundamental PITCHT listeners; overt. PITCHT = only overtone PITCHT listeners.

Variable	**PITCHT**		**fund. PITCHT**		**overt. PITCHT**	
	r	***p***	***r***	***p***	***r***	***p***
N1 latency	-0.0140	0.9392 n.s.	0.0120	0.9625 n.s.	-0.1685	0.5647 n.s.
N1 amplitude	0.2964	0.0995 n.s.	0.3290	0.1825 n.s.	-0.4659	0.0932 n.s.
P2 latency	-0.2303	0.2047 n.s.	-0.4890	**0.0394***	0.2362	0.4162 n.s.
P2 amplitude	-0.4200	**0.0167***	-0.3116	0.2081 n.s.	0.0192	0.9480 n.s.

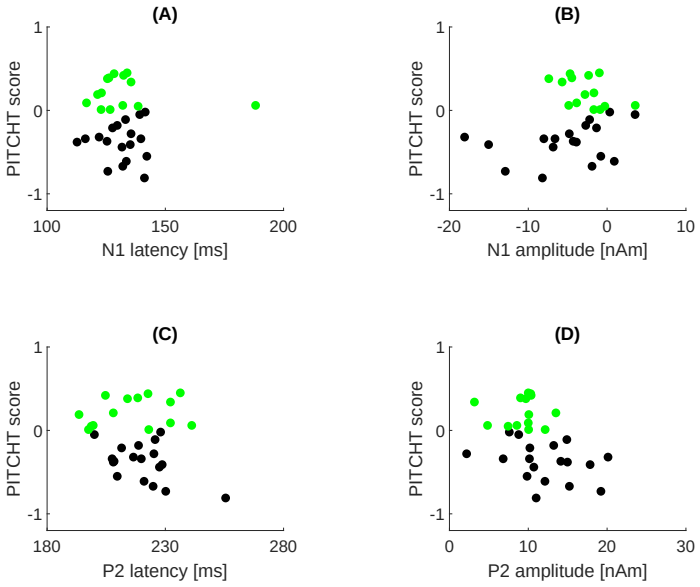


Figure 31: Combined display of the neuromagnetic data and PITCHT scores, both obtained from the NEURO subject group. The neuromagnetic data comprised the latencies and amplitudes of the subjects' individual N1 and P2 responses across the 12 instrumental tones and are represented on the x-axis in [ms] (latency) and [nAm] (amplitude). The y-axis shows the subjects' individual PITCHT scores as values between -1 and 1. Each dot represents the data obtained from one of the 32 NEURO subjects. (**A**) and (**B**) display the PITCHT score in relation with the N1 latency and N1 amplitude, respectively; (**C**) and (**D**) depict the PITCHT score in relation with the P2 latency and P2 amplitude, respectively. Note that *larger* N1 responses have *lower* amplitude values. In each subplot, black dots indicate fundamental PITCHT listeners and green dots indicate overtone PITCHT listeners.

3.5.3 Neuromagnetic Data and Sound Properties

To find out how the neuromagnetic data and the physical sound properties were related with each other, correlation analyses were performed between the subjects' N1 and P2 latencies/amplitudes and the instrumental tones' periodicity, hissiness, attack time and spectral centroid. It was found that the N1 latency was significantly correlated with the periodicity, attack time and spectral centroid (table 21): the N1 latency significantly increased with rising sound periodicity (fig. 32 **A**) and rising attack time (fig. 34 **A**), and significantly decreased with rising spectral centroid (fig. 35 **A**). Low register tones (i.e. bassoon tones and bass clarinet tones) exhibited longer periodicities (fig. 32 **A**) and attack times (fig. 34 **A**), and produced N1 responses with longer latencies than high register tones (i.e. clarinet tones and oboe tones) (figures 32–34 **A**); the spectral centroid was higher in high register tones and lower in low register tones (fig. 35 **A**). The N1 amplitude was significantly larger for sounds with longer periodicities (bassoon tones, bass clarinet tones) than for sounds with shorter periodicities (oboe tones, clarinet tones) (table 21, fig. 32 **B**). Moreover, the N1 amplitude significantly increased with decreasing hissiness (table 21, fig. 33 **B**). Hissiness tended to be smaller for low register tones than for high register tones (fig. 33). Concerning the P2 latency, significant correlations were found for the periodicity, attack times and spectral centroid (table 21): like the N1 latency, the P2 latency significantly increased with rising periodicity (fig. 32 **C**) and rising attack time (fig. 34 **C**), and significantly decreased with rising spectral centroid (fig. 35 **C**). Again, the latency of the neuromagnetic (P2) response was longer for the low register tones than for the high register tones. Other than the N1 amplitude, the P2 amplitude was significantly larger for tones with shorter periodicities (fig. 32 **D**) or greater hissiness (fig. 33 **D**) than for tones with longer periodicities or smaller hissiness (table 21). Beyond this, the P2 amplitude significantly rose with increasing spectral centroid (fig. 35 **D**).

Table 21: Results of the correlation tests between the neuromagnetic data obtained from the NEURO subject group and sound properties which were either extracted from the stabilised auditory images (hissiness, periodicity) or analysed by Christoph Reuter (University of Vienna, Vienna) using PADMEA (attack time, spectral centroid). The correlation tests were performed between the subjects' average N1 and P2 latencies and amplitudes for each of the 12 instrumental tones and each tone's hissiness, periodicity, attack time and spectral centroid.

Variable	**N1 latency**		**N1 amplitude**		**P2 latency**		**P2 amplitude**	
	r	***p***	***r***	***p***	***r***	***p***	***r***	***p***
Periodicity	0.9204	$<$ **0.001*** **	-0.59162	**0.0427***	0.8689	$<$ **0.001*** **	-0.7986	**0.002****
Hissiness	-0.3679	0.2394 n.s.	0.6141	**0.0336***	0.3864	0.2147 n.s.	0.6044	**0.0374***
Attack time	0.7502	**0.005****	0.0792	0.8067 n.s.	0.7046	**0.011***	-0.1432	0.6571 n.s.
Spectral centroid	-0.7592	**0.004****	0.5051	0.0940 n.s.	-0.7323	**0.007****	0.8735	$<$ **0.001*** **

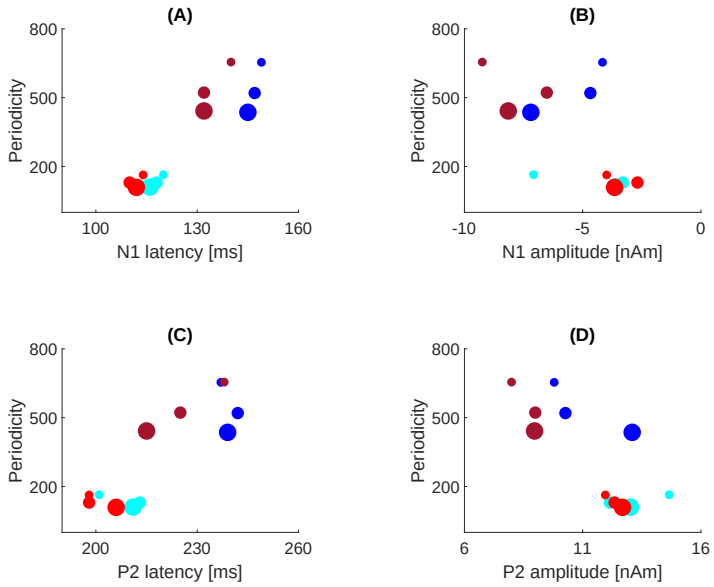


Figure 32: Combined display of the neuromagnetic data (obtained from the NEURO subject group) and the periodicity of each tone. The subjects' average N1/P2 latencies and amplitudes are represented on the x-axis in [ms] (latency) and in [nAm] (amplitude); the tones' individual periodicity is represented on the y-axis. (**A**) and (**B**) display the periodicity in relation with the N1 latency and N1 amplitude, respectively; (**C**) and (**D**) depict the periodicity in relation with the P2 latency and P2 amplitude, respectively. Note that *larger* N1 responses have *lower* amplitude values. In each subplot, dark blue dots represent the data for the bass clarinet tones D2 (small dot), F♯2 (middle-sized dot) and A2 (big dot); dark red dots depict the data for the bassoon tones D2 (small dot), F♯2 (middle-sized dot) and A2 (big dot); light blue dots represent the data for the clarinet tones D4 (small dot), F♯4 (middle-sized dot) and A4 (big dot), and light red dots depict the data for the oboe tones D4 (small dot), F♯4 (middle-sized dot) and A4 (big dot).

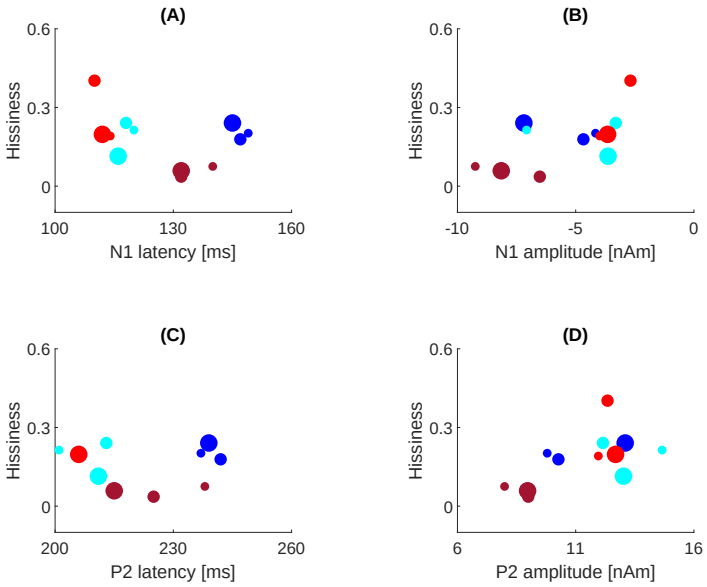


Figure 33: Combined display of the neuromagnetic data (obtained from the NEURO subject group) and the hissiness of each tone. The subjects' average N1/P2 latencies and amplitudes are represented on the x-axis in [ms] (latency) and in [nAm] (amplitude); the tones' individual hissiness is represented on the y-axis. (**A**) and (**B**) display the periodicity in relation with the N1 latency and N1 amplitude, respectively; (**C**) and (**D**) depict the periodicity in relation with the P2 latency and P2 amplitude, respectively. Note that *larger* N1 responses have *lower* amplitude values. In each subplot, dark blue dots represent the data for the bass clarinet tones D2 (small dot), F♯2 (middle-sized dot) and A2 (big dot); dark red dots depict the data for the bassoon tones D2 (small dot), F♯2 (middle-sized dot) and A2 (big dot); light blue dots represent the data for the clarinet tones D4 (small dot), F♯4 (middle-sized dot) and A4 (big dot), and light red dots depict the data for the oboe tones D4 (small dot), F♯4 (middle-sized dot) and A4 (big dot).

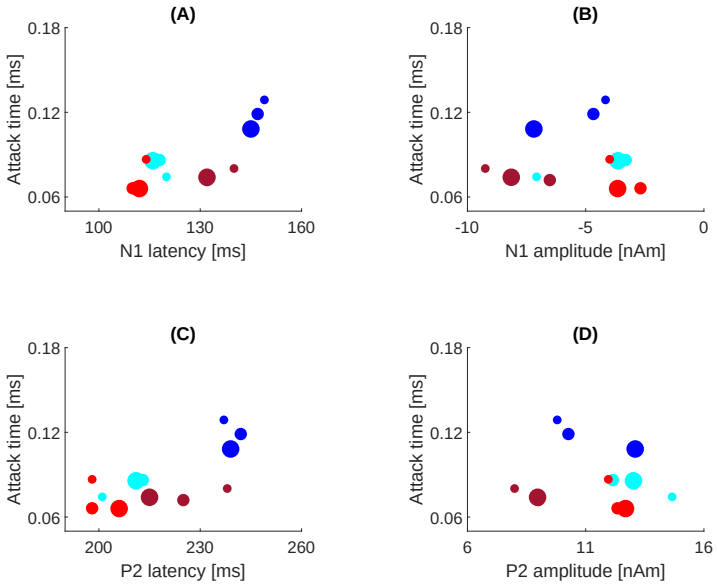


Figure 34: Combined display of the neuromagnetic data (obtained from the NEURO subject group) and the attack time of each tone. The subjects' average N1/P2 latencies and amplitudes are represented on the x-axis in [ms] (latency) and in [nAm] (amplitude); the tones' individual attack time is represented on the y-axis in [ms]. (**A**) and (**B**) display the attack time in relation with the N1 latency and N1 amplitude, respectively;(**C**) and (**D**) depict the attack time in relation with the P2 latency and P2 amplitude, respectively. Note that *larger* N1 responses have *lower* amplitude values. In each subplot, dark blue dots represent the data for the bass clarinet tones D2 (small dot), F♯2 (middle-sized dot) and A2 (big dot); dark red dots depict the data for the bassoon tones D2 (small dot), F♯2 (middle-sized dot) and A2 (big dot); light blue dots represent the data for the clarinet tones D4 (small dot), F♯4 (middle-sized dot) and A4 (big dot), and light red dots depict the data for the oboe tones D4 (small dot), F♯4 (middle-sized dot) and A4 (big dot).

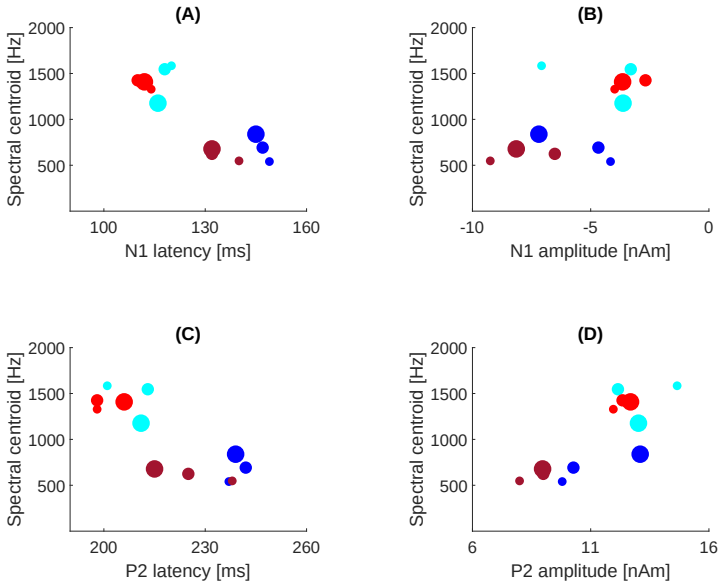


Figure 35: Combined display of the neuromagnetic data (obtained from the NEURO subject group) and the spectral centroid of each tone. The subjects' average N1/P2 latencies and amplitudes are represented on the x-axis in [ms] (latency) and in [nAm] (amplitude); the tones' individual spectral centroid is represented on the y-axis in [Hz]. (**A**) and (**B**) display the spectral centroid in relation with the N1 latency and N1 amplitude, respectively; (**C**) and (**D**) depict the spectral centroid in relation with the P2 latency and P2 amplitude, respectively. Note that *larger* N1 responses have *lower* amplitude values. In each subplot, dark blue dots represent the data for the bass clarinet tones D2 (small dot), F♯2 (middle-sized dot) and A2 (big dot); dark red dots depict the data for the bassoon tones D2 (small dot), F♯2 (middle-sized dot) and A2 (big dot); light blue dots represent the data for the clarinet tones D4 (small dot), F♯4 (middle-sized dot) and A4 (big dot), and light red dots depict the data for the oboe tones D4 (small dot), F♯4 (middle-sized dot) and A4 (big dot).

3.5.4 Psychoacoustic Data and Sound Properties

With the intention of finding out how the psychoacoustic data and physical sound data were related with each other, correlation tests were performed between the subject's brightness judgements and the instrumental tones' periodicity, hissiness, attack time and spectral centroid. The results reveal significant correlations between relative brightness and periodicity, as well as between relative brightness and spectral centroid (table 22): the relative brightness increased significantly with decreasing periodicity (fig. 36) and rising spectral centroid (fig. 37). The high register tones with the shorter periodicities and higher spectral centroids were judged brighter than the low register tones with the longer periodicities and lower spectral centroids.

Table 22: Results of the correlation tests between the psychoacoustic data obtained from the PSYCH subject group and sound properties which were either extracted from the stabilised auditory images (hissiness, periodicity) or analysed by Christoph Reuter (University of Vienna, Vienna) using PADMEA (attack time, spectral centroid). The correlation tests were performed between the subjects' average brightness judgements for each of the 12 instrumental tones and each tone's hissiness, periodicity, attack time and spectral centroid.

Variable	r	p
Periodicity	-0.9729	$< \mathbf{0.001}^{***}$
Hissiness	0.4282	0.1650 n.s.
Attack time	-0.5374	0.0716 n.s.
Spectral centroid	0.8325	$< \mathbf{0.001}^{***}$

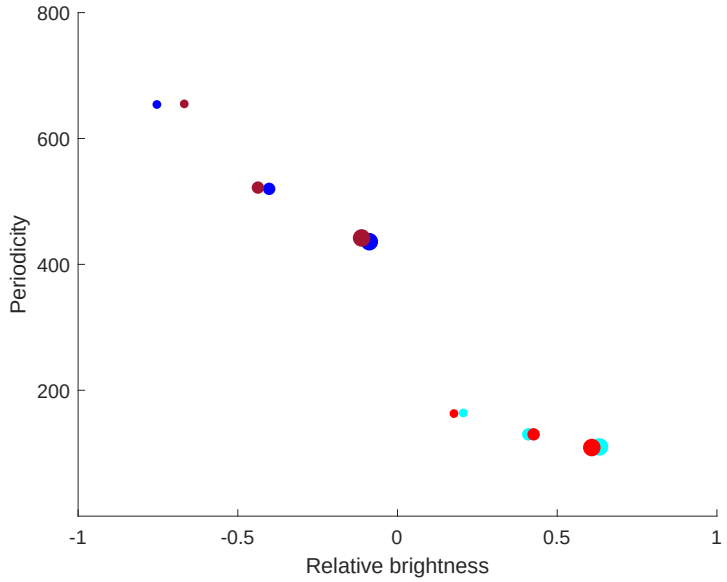


Figure 36: Combined display of the psychoacoustic data (obtained from the PSYCH subject group) and the periodicity of each tone. The subjects' average brightness judgements are represented on the x-axis as a value between -1 and 1; the tones' individual periodicity is represented on the y-axis. Dark blue dots represent the data for the bass clarinet tones D2 (small dot), F♯2 (middle-sized dot) and A2 (big dot); dark red dots depict the data for the bassoon tones D2 (small dot), F♯2 (middle-sized dot) and A2 (big dot); light blue dots represent the data for the clarinet tones D4 (small dot), F♯4 (middle-sized dot) and A4 (big dot), and light red dots depict the data for the oboe tones D4 (small dot), F♯4 (middle-sized dot) and A4 (big dot).

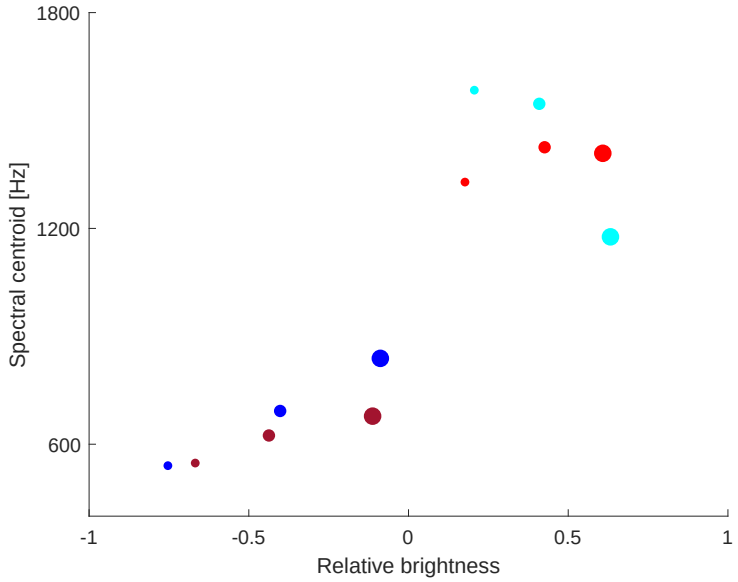


Figure 37: Combined display of the psychoacoustic data (obtained from the PSYCH subject group) and the spectral centroid of each tone. The subjects' average brightness judgements are represented on the x-axis as a value between -1 and 1; the tones' individual spectral centroid is represented on the y-axis in [Hz]. Dark blue dots represent the data for the bass clarinet tones D2 (small dot), F♯2 (middle-sized dot) and A2 (big dot); dark red dots depict the data for the bassoon tones D2 (small dot), F♯2 (middle-sized dot) and A2 (big dot); light blue dots represent the data for the clarinet tones D4 (small dot), F♯4 (middle-sized dot) and A4 (big dot), and light red dots depict the data for the oboe tones D4 (small dot), F♯4 (middle-sized dot) and A4 (big dot).

3.5.5 Psychoacoustic Data, AMMA Scores and PITCHT Scores

In order to exclude AMMA and PITCHT-related effects on the subjects' brightness judgements, correlation tests were performed between the subjects' individual AMMA scores/PITCHT scores and brightness judgements, averaged across the 12 instrumental tones. Results of the correlation tests are given in table 23; no significant correlations were found, neither between the relative brightness and the AMMA score, nor between the relative brightness and the PITCHT score. This was observed for all listeners, as well as for the subgroups of low/high AMMA listeners and fundamental/overtone PITCHT listeners. Figures 38 and 39 show that the relative brightness values were more similar between the subjects than the AMMA scores and PITCHT scores.

Table 23: Results of the correlation tests between the psychoacoustic data and AMMA scores, and between the psychoacoustic data and PITCHT scores. All data were obtained from the PSYCH subject group. The correlation tests were performed between the subjects' individual AMMA scores and brightness judgements, and between the subjects' individual PITCHT scores and brightness judgements. In both cases, the scores and brightness judgements were averaged across the 12 instrumental tones.AMMA/PITCHT = all listeners; low AMMA = low AMMA listeners, high AMMA = high AMMA listeners; fundamental PITCHT = fundamental PITCHT listeners, overtone PITCHT = overtone PITCHT listeners.

Variable	r	p
AMMA	-0.0084	0.9746 n.s.
- low AMMA	-0.2478	0.5921 n.s.
- high AMMA	-0.3994	0.2528 n.s.
PITCHT	-0.3305	0.1950 n.s.
- fundamental PITCHT	-0.0572	0.8752 n.s.
- overtone PITCHT	0.1295	0.7820 n.s.

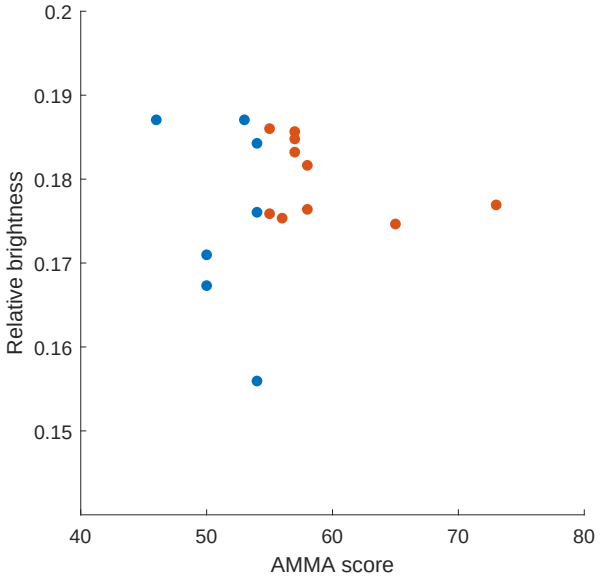


Figure 38: Combined display of the psychoacoustic data and AMMA scores, both obtained from the PSYCH subject group. The subjects' individual AMMA scores are represented on the x-axis in absolute numbers. The subjects' individual brightness judgements are represented on the y-axis as a relative brightness value between -1 and 1. Each dot represents the data obtained from one of the 17 PSYCH subjects. Blue dots indicate low AMMA listeners and red dots indicate high AMMA listeners.

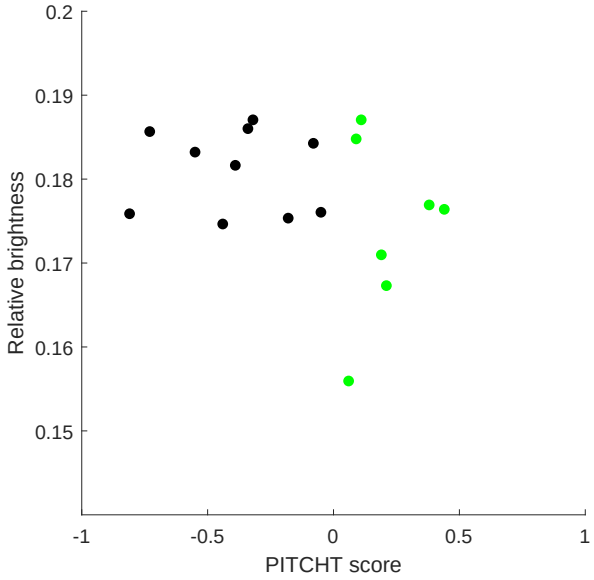


Figure 39: Combined display of the psychoacoustic data and PITCHT scores, both obtained from the PSYCH subject group. The subjects' individual PITCHT scores are represented on the x-axis as values between -1 and 1. The subjects' individual brightness judgements are represented on the y-axis as a relative brightness value between -1 and 1. Each dot represents the data obtained from one of the 17 PSYCH subjects. Black dots indicate fundamental PITCHT listeners and green dots indicate overtone PITCHT listeners.

3.5.6 Results of Factor Analysis

The factor analysis was performed on the subjects' average N1 and P2 latencies/amplitudes, the subjects' average brightness judgements, and the tones' individual periodicity, hissiness, attack time and spectral centroid. The rationale behind the factor analysis was to find out which principal components (factors) explained most of the variance in the data and can thus help to characterise the different dimensions of timbre. In total, nine factors were found in the principal component analysis, and the first two factors explained most of the variance in the data (eigenvalue ≥ 1; fig. 40). Table 24 summarises the rotated factor loadings and the communality values for the investigated variables. The communality values indicate that all variables were well represented by the two factors, and they show that the periodicity, N1 latency and attack time explained more than 90 % of the data variance. Variables loading on factor 1 included the attack time, N1 latency, P2 latency, periodicity and relative brightness; variables loading on factor 2 comprised the hissiness, N1 amplitude, P2 amplitude, spectral centroid and periodicity. Particularly high positive loadings on factor 1 were observed for the attack time and N1 latency; the hissiness was highly positively loaded on factor 2. These findings indicate that factor 1 primarily describes temporal sound attributes and temporal characteristics of the neuromagnetic responses; factor 2 rather describes spectral sound attributes and size characteristics of the neuromagnetic responses. In sum, factor 1 explained more variance than factor 2.

Table 24: Results of the factor analysis after data rotation using the VARIMAX method. For each variable, the rotated factor loadings and the communality are given.

Variable	**Factor 1**	**Factor 2**	**Communality**
Attack time	0.9280	0.2380	0.9177
N1 latency	0.9007	-0.3799	0.9556
P2 latency	0.8694	-0.3597	0.8853
Periodicity	0.7664	-0.6237	0.9764
Relative brightness	-0.7561	0.5696	0.8961
Hissiness	-0.0372	0.8540	0.7307
N1 amplitude	-0.1187	0.7987	0.6519
P2 amplitude	-0.3907	0.7724	0.7492
Spectral centroid	-0.5908	0.7141	0.8590
Explained variance	4.1045	3.5176	

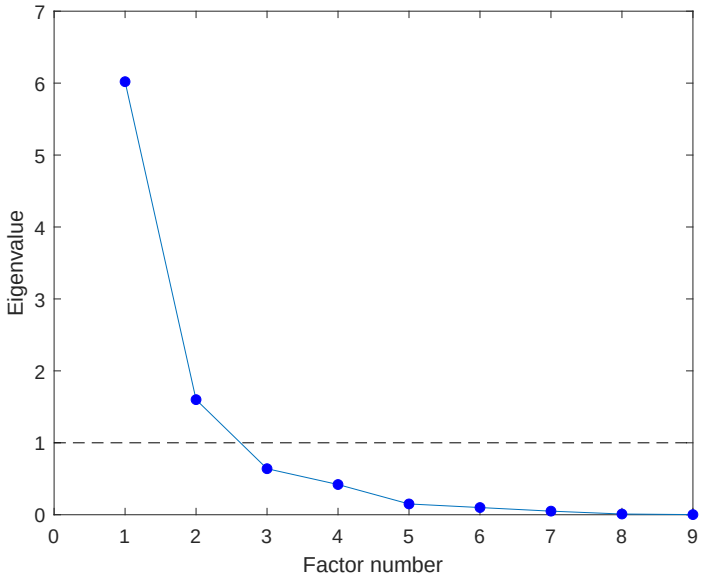


Figure 40: Results of the principal component analysis. The scree plot depicts the eigenvalue (y-axis) for each of the nine found factors 1–9 (x-axis). The eigenvalues indicate to which extent the individual factors could explain the observed variance in the data; the higher the eigenvalue, the more variance was explained by the respective factor. Factors with an eigenvalue < 1 were considered as factors of minor importance, whereas factors with an eigenvalue ≥ 1 were considered as factors of major importance. The cut-off eigenvalue 1 is represented by the black dashed line.

4 Discussion

4.1 Discussion of the Study Results

The objectives of this study included **(1)** analysing the neuromagnetic responses to changes in pitch contour, instrument register, instrument family and sound location and identifying their neural sources, **(2)** exploring the effect of natural and unified attack times on timbre processing, **(3)** investigating the relation between the neuromagnetic activity and the perceived brightness, **(4)** examining the relation between the neuromagnetic activity and physical sound properties, **(5)** assessing the relation between physical sound properties and the perceived brightness, **(6)** analysing the effects of musical aptitude, musical experience and pitch perception type on timbre processing. The overall aim of this project was to bring three different research areas – namely neuroscience, musicology and physics – together in order to improve the understanding of timbre processing and representation in the human auditory cortex.

Concerning the neuromagnetic responses, it was expected that melodies with changes in pitch contour/instrument register/instrument family/sound location evoke stronger and earlier N1 and P2 responses than melodies without these changes. However, it was found that melodies with register change influence the N1 and P2 responses more strongly than melodies with contour, family or sound location change. This result is in line with previous studies (e.g. Andermann et al., 2011, Andermann et al., 2017) which already indicated that register changes largely impact the latency and amplitude of neuromagnetic responses; furthermore, as these studies focus – among others – on voice sounds, it can be concluded that register-related latency and amplitude effects are produced by both instrumental stimuli and speech stimuli. In total, changes in instrument register produced the largest $N1_{cng}$ and $P2_{cng}$responses, whereas changes in pitch contour evoked the smallest $N1_{cng}$ and $P2_{cng}$ responses (fig. 14). A possible reason for this might be that changes in instrument register include especially large f_0 shifts which are more salient and therefore produce larger N1 and P2 responses (Andermann et al., 2021; Quiroga-Martinez et al., 2020; Regev et al., 2021) than the other changes which include smaller (and less salient) f_0 shifts. Probably, changes in pitch contour produced the smallest responses because melodies with contour

change contained f_0 shifts at all tone transitions and thus, the f_0 shift at the transition from the third to the fourth tone was not a special event – in contrast to the register/family/sound location change which only occurred once and only at the transition from the third to the fourth tone of the melody.

In several cases, significant N1/P2 latency and/or amplitude differences were observed: **(1)** melodies *without* contour change evoked significantly earlier $N1_{post}$ responses than melodies with contour change, but melodies *with* contour change elicited significantly larger $P2_{post}$ responses than melodies without contour change; **(2)** melodies *with* register change produced significantly earlier and larger $N1_{cng}$ responses and significantly larger $P2_{cng}$ responses than melodies without register change; **(3)** melodies *without* family change evoked significantly earlier $P2_{cng}$ responses than melodies with family change; **(4)** melodies *without* sound location change produced significantly earlier $N1_{cng}$ responses and significantly larger $P2_{cng}$ responses than melodies with sound location change. In the remaining cases, no significant N1/P2 latency and/or amplitude differences were observed. On the one hand, the results might indicate that changes in pitch contour, instrument family and sound location are less salient than changes in instrument register; on the other hand, the observation that melodies with contour change elicited significantly stronger $P2_{post}$ responses than melodies without contour change, does not fit into this picture, and it is rather improbable that the comparatively small pitch contour changes are more salient than family changes and sound location changes. By contrast, it is more likely that the contour, family and sound location changes *appear* less influential because the applied paradigm did not include melodies without any change, i.e. melodies with a respective change (e.g. contour) were contrasted to melodies which did not contain the respective change (i.e. contour) but the respective other three changes (i.e. register, family, sound location). Thus, as the register main effect was larger than the other main effects, the contour, family and sound effects might have lost significance when comparing them with the register effect.

Melodies with unified attack times elicited significantly stronger and earlier $N1_{on}$ responses and significantly larger $P2_{on}$ responses than melodies with natural attack times. This result was expected and is perfectly in line with studies showing that tones with more temporal regularity produce higher neuromagnetic activity than tones with less temporal regularity (e.g. Kim et al., 2022). Usually, the attack time is an important acoustic clue for identifying and differentiating musical instruments. In melodies made of unchanged (i.e. *natural*) tones, the attack time varied from tone to tone; by contrast, in melodies built from standardised (i.e. *unified*) tones, the attack time was identical for all tones. Consequently, in the *natural* condition,

the different attack times could be used for instrument identification; in the *unified* condition, the attack times lacked instrument-specific information and therefore, other acoustic features needed to be applied for instrument identification. Maybe in this case, the auditory cortex mainly applied spectral sound features (such as e.g. spectral centroid or formants) because they contained more instrument-specific information than the attack time. In fact, studies show that the identification of non-percussive sounds does not rely on the attack time, but rather on other sound parts such as the steady state (Suied et al., 2014).

The observation of a hemisphere effect with significantly shorter $N1_{cng}$ latency in the right hemisphere fits well with the finding that the right hemisphere is more involved in timbre processing than the left hemisphere (e.g. Formisano et al., 2008; Town and Bizley, 2013). However, there was no hemisphere effect for the $N1_{cng}$ amplitude, the $P2_{cng}$ latency and $P2_{cng}$ amplitude. Maybe this could indicate that the right hemisphere is rather implicated in early timbre processing, but subsequent processes increasingly require the activity of both hemispheres. Studies found that the left hemisphere is rather involved in the processing of distinct acoustic features and it was shown that especially posterior regions of the left auditory cortex are activated during timbre processing (Deike et al., 2004; Town and Bizley, 2013). In combination with the supposition that timbre might be processed along a hierarchy from the HG via the PT to the STG (Kumar et al., 2007), it is thinkable that the left hemisphere is rather involved in later timbre processing steps, whereas the right hemisphere is more involved in earlier timbre processing steps.

Another important result is that subjects with high musical aptitude exhibited significantly stronger P2 responses than subjects with low musical aptitude. Previous studies already indicated that high AMMA listeners are more sensitive towards and react more strongly to pitch and pitch changes than low AMMA listeners (e.g. Andermann et al., 2021). Thus, the herein obtained results might suggest that high AMMA listeners are also more sensitive to timbre and timbre changes than low AMMA listeners. Furthermore, previous studies showed that musical aptitude largely influences the N1 response but not the P2 response (e.g. Andermann et al., 2021). In the current study, the opposite was observed: musical aptitude largely influenced the $P2_{cng}$ response but not the $N1_{cng}$ response. Maybe this result highlights the difference between pitch and timbre processing: the N1 response might primarily represent pitch processing, whereas the P2 response might rather represent timbre processing. This would also fit with other research results which indicate that the P2 response reflects the processing of spectral sound cues (e.g. Kaneshiro et al., 2020). Significant interactions between the N1/P2 amplitudes and the AMMA score

show that some of the observed contour, register and sound location main effects are more pronounced in high AMMA listeners than in low AMMA listeners. More specifically, high AMMA listeners exhibited larger amplitude differences regarding the $N1_{post}$ response to melodies with/without contour change, concerning the $P2_{cng}$ response to melodies with/without register change, and regarding the $P2_{cng}$ response to melodies with/without sound location change. Thus, high AMMA listeners seem to be more sensitive to melodies *without* contour change, melodies *with* register change, and melodies *without* sound location change than low AMMA listeners. This implies that the individual musical aptitude does not only affect pitch processing (e.g. Andermann et al., 2021) but also timbre processing, and high musical aptitude might increase the sensitivity to instrument register (changes).

When comparing the neuromagnetic responses of fundamental PITCHT and overtone PITCHT listeners, it was found that fundamental PITCHT listeners exhibited significantly larger $N1_{cng}$ and $P2_{cng}$ responses than overtone PITCHT listeners. Perhaps fundamental PITCHT listeners react more strongly to timbre changes due to an advantage in brightness processing: brightness is correlated with the position of the spectral centroid (Peeters, 2004) and the position of the spectral centroid largely depends on the fundamental frequency (Czedik-Eysenberg, 2016, pg. 34). Thus, it might be easier for fundamental PITCHT listeners to use the formant frequency for sound identification. Significant interactions between the N1/P2 amplitudes and the PITCHT score reveal that several contour, register, family and sound location main effects are more pronounced in fundamental PITCHT listeners than in overtone PITCHT listeners. In detail, fundamental PITCHT listeners presented larger amplitude differences regarding the $P2_{post}$ response to melodies with/without contour change, concerning the $N1_{cng}$ response to melodies with/without register change, regarding the $P2_{cng}$ response to melodies with/without family change, and in relation to the $P2_{cng}$ response to melodies with/without sound location change. In total, fundamental PITCHT listeners seem to be more sensitive to melodies *with* contour change, melodies *with* register change, melodies *without* family change, melodies *with* sound location change than overtone PITCHT listeners. This finding suggests that the individual pitch perception type may affect timbre processing, and subjects with lower PITCHT scores might be more sensitive towards pitch contour, instrument register and sound location than subjects with higher PITCHT scores.

Concerning the neural source locations, it was observed that the $N1_{post}$ sources for melodies with contour change are located significantly anterior in the HG than the $N1_{cng}$ sources for melodies with register/family/sound location change. This result matches the expectation as pitch changes are primarily processed in the

anterior part of the HG. Interestingly, the $N1_{cng}$ and $P2_{cng}$ sources for melodies with sound location change were located significantly more cranial than the $N1_{cng}$ and $P2_{cng}$ sources for the other melodies. Additionally, substantial differences could be observed between the left and right hemispheres. Both the $N1_{cng}$ and the $P2_{cng}$ sources were located more closely in the left hemisphere and more distributed in the right hemisphere, and the posterior-anterior sequence of the sources varied between the hemispheres. However, there were some exceptions: in both hemispheres, the $N1_{post}$ sources for melodies with contour change were the most anterior sources and the $P2_{post}$ sources for melodies with contour change were the second most anterior sources. Furthermore, the $N1_{cng}$ sources for melodies with sound location change were the second most posterior sources and the $P2_{cng}$ sources for melodies with sound location change were the most posterior sources. In sum, the N1 sources were located more anterior in the two hemispheres than expected. Taken together, these findings suggest that the processing of pitch contour and sound location might be represented more similarly in the two hemispheres than the processing of instrument register and instrument family. Conversely, this would mean that instrument register and instrument family are processed differently in the two hemispheres, which would be in line with the hypothesis that the right and left hemispheres bring in different specialisations for timbre processing. More specifically, the obtained data indicate that instrument family is processed more anteriorly in the left hemisphere than in the right hemisphere, whereas instrument register is processed more anteriorly in the right hemisphere than in the left hemisphere. Another interesting finding is that in the right hemisphere, all N1 and P2 sources were located in the HG, whereas in the left hemisphere, the N1 and P2 sources for melodies with sound location change were located at the border between PT and HG; this again supports the hypothesis that the left and right hemispheres are differentially involved in timbre processing.

The results of the psychoacoustic experiment show that tones of high register instruments (oboe, clarinet) were perceived brighter than tones of low register instruments (bassoon, bass clarinet). This is not surprising as pitch has a substantial influence on brightness perception and high pitched tones are known to sound brighter than low pitched tones. Consequently, it was not unusual either that subjects judged the highest oboe tone (A4) and the highest clarinet tone (A4) as brightest. For each of the four instruments, the perceived brightness differed significantly between the three tones D, F♯ and A. This finding is in line with previous studies and highlights that pitch and brightness perception indeed interact. It was also found that the subjects' brightness judgements differed between the four instruments: the perceived brightness differed significantly between tones that varied in register, register & pitch, register & family, or register & family & pitch; however,

brightness judgements did not differ significantly between tones varying only in instrument family. Together, these findings might indicate that brightness is more influenced by register and pitch than by the instrument as such – which matches previous research results demonstrating that timbre primarily depends on acoustic sound features and less on the type of instrument (Siddiq et al., 2015; Reuter, 2019).

The combined analyses on the neuromagnetic data, psychoacoustic data and sound properties revealed further important findings: **(1)** perceptually brighter tones evoked significantly earlier N1 responses and significantly larger P2 responses than perceptually duller tones; **(2)** tones with shorter periodicities, shorter attack times or higher spectral centroids elicited significantly earlier N1 and P2 responses than tones with longer periodicities, longer attack times or lower spectral centroids; **(3)** tones with longer periodicities or smaller hissiness produced significantly stronger N1 responses than tones with shorter periodicities or larger hissiness; by contrast, **(4)** tones with shorter periodicities, smaller hissiness or higher spectral centroids evoked larger P2 responses than tones with longer periodicities or larger hissiness; **(5)** tones with shorter periodicities or higher spectral centroids were perceived significantly brighter than tones with longer periodicities or lower spectral centroids. As higher pitched sounds tend to have shorter periodicities, shorter attack times and higher spectral centroids than lower pitched sounds, and the applied high register tones (i.e. oboe tones, clarinet tones) were higher in pitch than the presented low register tones (i.e. bassoon tones, bass clarinet tones), it can be concluded that the high register tones were perceived brighter and produced earlier N1 responses as well as earlier and larger P2 responses than the low register tones; in contrast to this, the low register tones evoked larger N1 responses than the high register tones. A possible explanation for this phenomenon could be that the N1 response might be more influenced by temporal sound aspects (such as periodicity and attack time), whereas the P2 response might be more influenced by spectral sound aspects (such as hissiness and spectral centroid). This is in line with a study by Rupp et al. (2013) who showed that the hissiness of a sound is strongly correlated with the evoked P2 amplitude.

Further important results were obtained from the combined analyses on the subjects' *individual* neuromagnetic/psychoacoustic data, AMMA scores and PITCHT scores: **(1)** the N1 amplitude significantly increased with rising AMMA score; **(2)** the P2 amplitude significantly increased with decreasing PITCHT score; **(3)** the P2 latency became significantly shorter with increasing PITCHT score, but only in the group of fundamental PITCHT listeners; **(4)** no significant correlations were found between the psychoacoustic data and the AMMA or PITCHT scores. These findings indicate that there were subject group-related effects on the neuromagnetic responses but not on the brightness judgements.

The factor analysis on the neuromagnetic data, psychoacoustic data and sound properties revealed two principal factors which together explain the vast majority of the data variance. It was found that the first factor mainly describes temporal sound attributes (attack time, periodicity) and temporal characteristics of the N1 and P2 responses (N1 latency, P2 latency); by contrast, the second factor primarily describes spectral sound attributes (hissiness, spectral centroid) and size characteristics of the N1 and P2 responses (N1 amplitude, P2 amplitude). As the first factor explains more variance than the second factor, temporal sound attributes and temporal response characteristics might influence timbre perception more than spectral sound attributes and N1 and P2 size characteristics.

Regarding the AMMA score, PITCHT score and practical musical experience, only minor differences were observed between the two subject groups NEURO and PSYCH: in both groups, the AMMA score rose significantly with increasing musical training duration and general musical activity and there was no significant correlation between the AMMA score and the age at musical training onset. A significant positive correlation between the AMMA score and musical training duration was already observed in a previous study (Andermann et al., 2021) and other investigations indicate that musical training improves auditory skills and thereby increases individual musicality (e.g. Hyde et al., 2009). The most striking observation was that the PITCHT scores differed significantly between male and female subjects of the NEURO group. This effect is very unusual and, as it was not observed in the PSYCH subject group, it is assumed that it was rather an exceptional coincidence. In fact, the data of the NEURO group show that the four most extreme scores were obtained by male subjects and in all four cases, the scores were close to -1. With respect to their composition, both the NEURO group and the PSYCH group comprised more high AMMA listeners than low AMMA listeners, and more fundamental PITCHT listeners than overtone PITCHT listeners. The largest difference between the two subject groups was that the NEURO group mainly comprised musically active persons, whereas the PSYCH group primarily included musically inactive persons.

4.2 Conclusion and Outlook

Taken together, this study revealed the following: **(1)** Changes in instrument register produce larger $N1_{cng}$ and $P2_{cng}$ responses than changes in pitch contour, instrument family or sound location – probably because register changes include larger pitch shifts which are more salient. **(2)** Contour changes and register changes have a stronger effect on the N1 and P2 latencies/amplitudes than family changes and sound location changes; contour changes largely impact the $P2_{post}$ amplitude, register changes mainly influence the $N1_{cng}$ latency, $N1_{cng}$ amplitude and $P2_{post}$ amplitude. **(3)** Standardised attack times elicit stronger $N1_{on}$ and $P2_{on}$ responses, presumably because standardised attack times contain less instrument-specific information and thus, the auditory cortex rather uses spectral sound features (instead of temporal sound features) for instrument identification. **(4)** The right auditory cortex seems to be more involved in earlier timbre processing, whereas the left auditory cortex seems to be more involved in later timbre processing. **(5)** Subjects with higher musical aptitude are more sensitive to *timbre* and timbre changes, probably because they are more musically experienced. This extends previous research results showing that subjects with higher musical aptitude are more sensitive to *pitch* and pitch changes. **(6)** Fundamental pitch listeners are more sensitive to timbre changes than overtone listeners; it is assumed that fundamental pitch listeners have an advantage in brightness processing due to the correlation between brightness, spectral centroid, and fundamental frequency. **(7)** Contour changes and sound location changes are more similarly represented in the two cerebral hemispheres than register changes and family changes; register changes are represented *more anterior in the right hemisphere* than in the left hemisphere, family changes are represented *more anterior in the left hemisphere than in the right hemisphere*. This supports the hypothesis that the right and left cerebral hemispheres are involved in different timbre processing tasks. **(8)** The perceived brightness of a tone is more influenced by pitch and the instrument register than by the instrument family. This is in line with previous studies demonstrating that timbre is more shaped by acoustic sound features than by the instrument type. **(9)** The N1 response is more influenced by temporal sound aspects (such as attack time and periodicity), whereas the P2 response is more influenced by spectral sound aspects (such as hissiness and spectral centroid). **(10)** Temporal sound attributes and temporal neuromagnetic response characteristics (i.e. N1 and P2 latencies) seem to have a stronger impact on timbre perception than spectral sound attributes and N1/ P2 size characteristics (i.e. N1 and P2 amplitudes). **(11)** Both the musical aptitude and pitch perception type have an impact on timbre processing.

In addition to these findings, there are several open questions which should be addressed in further studies. For instance, it is unclear why instrument register and instrument family are represented differently in the left and right hemispheres and how this correlates with timbre perception. Furthermore, it is striking that instrument family was found to be far less involved in timbre processing and timbre perception than pitch and instrument register. Although it is known that pitch and register have a greater influence on timbre than the instrument as such, it should be investigated whether the instrument family really *is* of such low importance or whether this effect was only due to the applied paradigm. Another open question is how in detail timbre is processed in the individual cerebral hemispheres and whether there really is a segregation regarding earlier and later timbre processing steps. It is interesting that all observed N1 sources were located more anterior than expected; thus, the origin of this effect should be examined in further timbre studies. A question of minor importance – but still of importance for better understanding the interaction between pitch and timbre processing – is why fundamental PITCHT listeners are more sensitive towards timbre (changes) than overtone PITCHT listeners. Thus, the interaction between pitch perception type and timbre processing should be addressed in more detail. Finally, the interactions between neuromagnetic responses, perceived brightness and physical sound properties should be studied more extensively and more physical sound properties should be included in the factor analysis to get a better overview which sound properties influence timbre perception and timbre processing.

5 Summary

The present work investigates the neuromagnetic representation of musical timbre dimensions in the human auditory cortex and aims at improving the understanding of timbre processing and perception. To this, neuromagnetic and psychoacoustic data were collected in a magnetoencephalography experiment and a psychoacoustic experiment, and the obtained data were set in relation to physical sound properties. Using magnetoencephalography, the subjects' neuromagnetic responses to melodies with changes in pitch contour/instrument register/instrument family/sound location, as well as neuromagnetic responses to melodies differing in attack time, were registered in order to investigate the neuromagnetic processing and representation of three established timbre dimensions (pitch, brightness and attack time). The melodies consisted of oboe, clarinet, bassoon and bass clarinet notes, which were judged regarding their relative brightness in the psychoacoustic experiment. Results show that contour changes largely influence the neuromagnetic P2 response to the fifth tone of the melody, and register changes mainly affect the neuromagnetic N1 and P2 responses to the fourth tone of the melody; by contrast, changes in pitch contour, instrument family and changes in sound location do not have a major influence on the N1 and P2 responses to the fourth tone of the melody. This suggests that register changes are more salient than pitch contour, instrument family and sound location changes. Standardised attack times evoked significantly larger onset-N1 and onset-P2 responses than natural attack times, which implies that standardised attack times contain less instrument-specific information such that the auditory cortex rather uses spectral than temporal sound features for instrument identification. Results further indicate that the right auditory cortex is more involved in early timbre processing, whereas the left auditory cortex is more implicated in later timbre processing. The perceived brightness of the instrumental tones was found to depend more on pitch and instrument register than on instrument family. This observation highlights that pitch and register have a stronger effect on timbre perception than instrument family. Temporal sound features are mainly reflected in the N1 response, whereas spectral sound features are largely mirrored in the P2 response. Timbre perception depends more on temporal sound attributes and N1/P2 latencies than on spectral sound attributes and N1/P2 amplitudes. Beyond that, timbre *processing* – but not timbre *perception* – is influenced by the subjects' musical aptitude and pitch perception preference.

6 Zusammenfassung

Die vorliegende Arbeit untersucht die neuromagnetische Repräsentation musikalischer Klangfarbendimensionen im menschlichen auditorischen Kortex und zielt darauf ab, das Verständnis der Klangfarbenverarbeitung und -wahrnehmung zu verbessern. Zu diesem Zweck wurden in einem Magnetenzephalographie-Experiment und einem psychoakustischen Experiment neuromagnetische und psychoakustische Daten erhoben, die anschließend in Bezug zu physikalischen Klangeigenschaften gesetzt wurden. Mit Hilfe der Magnetenzephalographie wurden die neuromagnetischen Reaktionen der Probanden in Bezug auf Melodien mit Änderungen in Tonhöhenkontur/Instrumentenregister/Instrumentenfamilie/Reizlokalisation, ebenso wie die Reaktionen auf Melodien mit unterschiedlicher Einschwingzeit, registriert, um die neuromagnetische Verarbeitung und Repräsentation drei etablierter Klangfarbendimensionen (Tonhöhe, Helligkeit und Einschwingzeit) zu untersuchen. Die Melodien bestanden aus Oboen-, Klarinetten-, Fagott- und Bassklarinettentönen, die im psychoakustischen Experiment hinsichtlich ihrer relativen Helligkeit beurteilt wurden. Ergebnisse der Studie zeigen, dass Konturänderungen weitgehend die neuromagnetische P2-Antwort auf den fünften Melodie-Ton beeinflussen und Registeränderungen hauptsächlich die neuromagnetischen N1- und P2-Antworten auf den vierten Melodie-Ton modulieren; im Gegensatz dazu haben Änderungen der Tonhöhenkontur, der Instrumentenfamilie und Änderungen der Reizlokalisation keinen bedeutsamen Einfluss auf die N1- und P2-Antworten bezüglich des vierten Melodie-Tons. Dies deutet darauf hin, dass Änderungen des Registers salienter sind als Änderungen der Tonhöhenkontur, Instrumentenfamilie oder Reizlokalisation. Standardisierte Einschwingzeiten riefen signifikant größere Onset-N1- und Onset-P2-Antworten hervor als natürliche Einschwingzeiten, was darauf schließen lässt, dass standardisierte Einschwingzeiten weniger instrumentenspezifische Informationen enthalten und der auditorische Kortex deshalb eher spektrale als zeitliche Klangmerkmale zur Instrumentenidentifikation heranzieht. Ergebnisse dieser Studie deuten ferner darauf hin, dass der rechte auditorische Kortex hauptsächlich an früheren Klangfarbenverarbeitungsschritten beteiligt ist, während der linke auditorische Kortex weitgehend an späteren Klangfarbenverarbeitungsschritten beteiligt zu sein scheint. Die wahrgenommene Helligkeit der Instrumentaltöne hing deutlich stärker von der Tonhöhe und dem Instrumentenregister ab als von der Instrumentenfamilie. Diese Beobachtung unterstreicht, dass Tonhöhe und Register einen größeren Einfluss auf die Klangfarbenwahrnehmung haben als die Instrumenten-

familie. Zeitliche Klangmerkmale spiegeln sich überwiegend in der N1-Antwort wider, während spektrale Klangmerkmale sich hauptsächlich in der P2-Antwort abzeichnen. Die Wahrnehmung der Klangfarbe hangt stärker von zeitlichen Klangmerkmalen und den N1/P2-Latenzen ab als von spektralen Klangmerkmalen und den N1/P2-Amplituden. Darüber hinaus wird die Klangfarben*verarbeitung* - nicht aber die Klangfarben*wahrnehmung* - durch die musikalische Begabung und Tonhöhenwahrnehmungspräferenz der Probanden beeinflusst.

7 Bibliography

Albersheim, G. (1939). *Zur Psychologie der Ton– und Klangeigenschaften*. Heitz und Co., Leipzig, Strassburg, Zuerich.

Allen, E., Burton, P., Mesik, J., Olman, C., & Oxenham, A. (2019). Cortical Correlates of Attention to Auditory Features. *J Neurosci*, *39*(17), 3292–3300. https://doi.org/10.1523/jneurosci.0588-18.2019

Allen, E., Burton, P., Olman, C., & Oxenham, A. (2016). Representations of Pitch and Timbre Variation in Human Auditory Cortex. *J Neurosci*, *37*(5), 1284–1293. https://doi.org/10.1523/jneurosci.2336-16.2016

Allen, E., & Oxenham, A. (2014). Symmetric interactions and interference between pitch and timbre. *J Acoust Soc Am*, *135*(3), 1371–1379. https://doi.org/10.1121/1.4863269

Andermann, M. (2014). *Einfluss von Phaseneffekten auf die Tonhöhenverarbeitung im auditorischen System* [Doctoral dissertation, Universitaet Heidelberg].

Andermann, M., Günther, M., Patterson, R., & Rupp, A. (2021). Early cortical processing of pitch height and the role of adaptation and musicality. *Neuroimage*, *225*, 117501. https://doi.org/10.1016/j.neuroimage.2020.117501

Andermann, M., Patterson, R., Vogt, C., Winterstetter, L., & Rupp, A. (2017). Neuromagnetic correlates of voice pitch, vowel type, and speaker size in auditory cortex. *Neuroimage*, *158*, 79–89. https://doi.org/10.1016/j.neuroimage.2017.06.065

Andermann, M., van Dinther, R., Patterson, R., & Rupp, A. (2011). Neuromagnetic representation of musical register information in human auditory cortex. *Neuroimage*, *57*(4), 1499–1506. https://doi.org/10.1016/j.neuroimage.2011.05.049

ANSI. (1960). *American Standard. Acoustical Terminology Including Mechanical Shock and Vibration*. Acoustical Society of America (ASA), New York.

Baillet, S. (2011). *Electromagnetic Brain Mapping Using MEG and EEG*. Oxford University Press, Oxford. https://doi.org/10.1093/oxfordhb/9780195342161.013.0007

Baillet, S. (2017). Magnetoencephalography for brain electrophysiology and imaging. *Nat Neurosci*, *20*(3), 327–339. https://doi.org/10.1038/nn.4504

Beal, A. (1985). The skill of recognizing musical structures. *Mem Cognit*, *13*(5), 405–412. https://doi.org/10.3758/bf03198453

Bear, M., Connors, B., & Paradiso, M. (2016). *Neurowissenschaften. Ein grundlegendes Lehrbuch für Biologie, Medizin und Psychologie* (A. Engel, Ed.; 3rd ed.). Springer, Berlin, Heidelberg.

Bendor, D., Osmanski, M., & Wang, X. (2012). Dual-Pitch Processing Mechanisms in Primate Auditory Cortex. *J Neurosci*, *32*(46), 16149–16161. https://doi.org/10.1523/jneurosci.2563-12.2012

Bendor, D., & Wang, X. (2005). The neuronal representation of pitch in primate auditory cortex. *Nature*, *436*(7054), 1161–1165. https://doi.org/10.1038/nature03867

Bendor, D., & Wang, X. (2006). Cortical representations of pitch in monkeys and humans. *Curr Opin Neurobiol*, *16*(4), 391–399. https://doi.org/10.1016/j.conb.2006.07.001

Bianchi, F., Hjortkjær, J., Santurette, S., Zatorre, R., Siebner, H., & Dau, T. (2017). Subcortical and cortical correlates of pitch discrimination: Evidence for two levels of neuroplasticity in musicians. *Neuroimage*, *163*, 398–412. https://doi.org/10.1016/j.neuroimage.2017.07.057

Bidelman, G., Krishnan, A., & Gandour, J. (2010). Enhanced brainstem encoding predicts musicians' perceptual advantages with pitch. *Eur J Neurosci*, *33*(3), 530–538. https://doi.org/10.1111/j.1460-9568.2010.07527.x

Bizley, J., & Cohen, Y. (2013). The what, where and how of auditory-object perception. *Nat Rev Neurosci*, *14*(10), 693–707. https://doi.org/10.1038/nrn3565

Bizley, J., Nodal, F., Parsons, C., & King, A. (2007). Role of Auditory Cortex in Sound Localization in the Midsagittal Plane. *J Neurophysiol*, *98*(3), 1763–1774. https://doi.org/10.1152/jn.00444.2007

Bizley, J., & Walker, K. (2010). Sensitivity and Selectivity of Neurons in Auditory Cortex to the Pitch, Timbre, and Location of Sounds. *Neuroscientist*, *16*(4), 453–469. https://doi.org/10.1177/1073858410371009

Bizley, J., Walker, K., Silverman, B., King, A., & Schnupp, J. (2009). Interdependent Encoding of Pitch, Timbre, and Spatial Location in Auditory Cortex. *J Neurosci*, *29*(7), 2064–2075. https://doi.org/10.1523/jneurosci.4755-08.2009

Boer, D., & Abubakar, A. (2014). Music listening in families and peer groups: Benefits for young people's social cohesion and emotional well-being across four cultures. *Front Psychol*, *5*. https://doi.org/10.3389/fpsyg.2014.00392

Bowling, D. (2023). Biological principles for music and mental health. *Transl Psychiatry*, *13*(1). https://doi.org/10.1038/s41398-023-02671-4

Bradley, R., & Terry, M. (1952). Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons. *Biometrika*, *39*(3/4), 324. https://doi.org/10.2307/2334029

Bregman, A. (1990). *Auditory Scene Analysis. The Perceptual Organization of Sound*. MIT Press, Cambridge.

Caclin, A., McAdams, S., Smith, B., & Winsberg, S. (2005). Acoustic correlates of timbre space dimensions: A confirmatory study using synthetic tones. *J Acoust Soc Am*, *118*(1), 471–482. https://doi.org/10.1121/1.1929229

Campbell, M., & Greated, C. (1994). *The musician's guide to acoustics*. Oxford University Press, Oxford.

Cariani, P. (1999). Temporal Coding of Periodicity Pitch in the Auditory System: An Overview. *Neural Plast*, *6*(4), 147–172. https://doi.org/10.1155/np.1999.147

Clark, M., Luce, D., Abrams, R., Schlossberg, H., & Rome, J. (1963). Preliminary Experiments on the Aural Significance of Parts of Tones of Orchestral Instruments and on Choral Tones. *J Audio Eng Soc*, *11*(1), 45–54. http://www.aes.org/e-lib/browse.cfm?elib=821

Clarke, S., Bellmann, A., De Ribaupierre, F., & Assal, G. (1996). Non-verbal auditory recognition in normal subjects and brain-damaged patients: Evidence for parallel processing. *Neuropsychologia*, *34*(6), 587–603. https://doi.org/10.1016/0028-3932(95)00142-5

Clarke, S., Bellmann, A., Meuli, R., Assal, G., & Steck, A. (2000). Auditory agnosia and auditory spatial deficits following left hemispheric lesions: Evidence for distinct processing pathways. *Neuropsychologia*, *38*(6), 797–807. https://doi.org/10.1016/s0028-3932(99)00141-4

Cohen, M., Grossberg, S., & Wyse, L. (1995). A spectral network model of pitch perception. *J Acoust Soc Am*, *98*(2), 862–879. https://doi.org/10.1121/1.413512

Czedik-Eysenberg, I. (2016). *Music Information Retrieval und Klangfarbe* [Master's thesis, Universitaet Wien].

Czedik-Eysenberg, I. (2021). *Semantische Modellierung wahrnehmungspsychologischer Musikdimensionen auf Basis von akustischen Signaleigenschaften* [Doctoral dissertation, Universitaet Wien].

Davis, S., & Mermelstein, P. (1980). Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *IEEE Trans Signal Process*, *28*(4), 357–366. https://doi.org/10.1109/tassp.1980.1163420

Deike, S., Gaschler-Markefski, B., Brechmann, A., & Scheich, H. (2004). Auditory stream segregation relying on timbre involves left auditory cortex. *Neuroreport*, *15*(9), 1511–1514. https://doi.org/10.1097/01.wnr.0000132919.12990.34

DiCiccio, T., & Efron, B. (1996). Bootstrap confidence intervals. *Stat Sci*, *11*(3). https://doi.org/10.1214/ss/1032280214

Elliott, T., Hamilton, L., & Theunissen, F. (2013). Acoustic structure of the five perceptual dimensions of timbre in orchestral instrument tones. *J Acoust Soc Am*, *133*(1), 389–404. https://doi.org/10.1121/1.4770244

Fitch, W. (1997). Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. *J Acoust Soc Am*, *102*(2), 1213–1222. https://doi.org/10.1121/1.421048

Fitch, W. (2006). The biology and evolution of music: A comparative perspective. *Cognition*, *100*(1), 173–215. https://doi.org/10.1016/j.cognition.2005.11.009

Formisano, E., De Martino, F., Bonte, M., & Goebel, R. (2008). "Who" Is Saying "What"? Brain-Based Decoding of Human Voice and Speech. *Science*, *322*(5903), 970–973. https://doi.org/10.1126/science.1164318

Fuller, B., Horii, Y., & Conner, D. (1992). Validity and reliability of nonverbal voice measures as indicators of stressor-provoked anxiety. *Res Nurs Health*, *15*(5), 379–389. https://doi.org/10.1002/nur.4770150507

Gelfer, M., & Mikos, V. (2005). The Relative Contributions of Speaking Fundamental Frequency and Formant Frequencies to Gender Identification Based on Isolated Vowels. *J Voice*, *19*(4), 544–554. https://doi.org/10.1016/j.jvoice.2004.10.006

Gordon, E. (1989). *Manual for the Advanced Measures of Music Audiation*. G.l.A. Publications Inc., Chicago.

Grey, J. (1977). Multidimensional perceptual scaling of musical timbres. *J Acoust Soc Am*, *61*(5), 1270–1277. https://doi.org/10.1121/1.381428

Grey, J., & Gordon, J. (1978). Perceptual effects of spectral modifications on musical timbres. *J Acoust Soc Am*, *63*(5), 1493–1500. https://doi.org/10.1121/1.381843

Grey, J., & Moorer, J. (1977). Perceptual evaluations of synthesized musical instrument tones. *J Acoust Soc Am*, *62*(2), 454–462. https://doi.org/10.1121/1.381508

Griffiths, T., Büchel, C., Frackowiak, R., & Patterson, R. (1998). Analysis of temporal structure in sound by the human brain. *Nat Neurosci*, *1*(5), 422–427. https://doi.org/10.1038/1637

Griffiths, T., & Green, G. (1999). Cortical Activation during Perception of a Rotating Wide-Field Acoustic Stimulus. *Neuroimage*, *10*(1), 84–90. https://doi.org/10.1006/nimg.1999.0464

Gutschalk, A., Patterson, R., Rupp, A., Uppenkamp, S., & Scherg, M. (2002). Sustained Magnetic Fields Reveal Separate Sites for Sound Level and Temporal Regularity in Human Auditory Cortex. *Neuroimage*, *15*(1), 207–216. https://doi.org/10.1006/nimg.2001.0949

Gutschalk, A., Patterson, R., Scherg, M., Uppenkamp, S., & Rupp, A. (2004). Temporal dynamics of pitch in human auditory cortex. *Neuroimage*, *22*(2), 755–766. https://doi.org/10.1016/j.neuroimage.2004.01.025

Heil, P. (1997a). Auditory Cortical Onset Responses Revisited. I. First-Spike Timing. *J Neurophysiol*, *77*(5), 2616–2641. https://doi.org/10.1152/jn.1997.77.5.2616

Heil, P. (1997b). Auditory Cortical Onset Responses Revisited. II. Response Strength. *J Neurophysiol*, *77*(5), 2642–2660. https://doi.org/10.1152/jn.1997.77.5.2642

Heil, P., Rajan, R., & Irvine, D. (1994). Topographic representation of tone intensity along the isofrequency axis of cat primary auditory cortex. *Hear Res*, *76*(1–2), 188–202. https://doi.org/10.1016/0378-5955(94)90099-x

Henderson, A. (2005). The bootstrap: A technique for data-driven statistics. using computer-intensive analyses to explore experimental data. *Clin Chim Acta*, *359*(1–2), 1–26. https://doi.org/10.1016/j.cccn.2005.04.002

Hyde, K., Lerch, J., Norton, A., Forgeard, M., Winner, E., Evans, A., & Schlaug, G. (2009). Musical Training Shapes Structural Brain Development. *J Neurosci*, *29*(10), 3019–3025. https://doi.org/10.1523/jneurosci.5118-08.2009

Irino, T., & Patterson, R. (2002). Segregating information about the size and shape of the vocal tract using a time-domain auditory model: The stabilised wavelet-Mellin transform. *Speech Commun*, *36*(3–4), 181–203. https://doi.org/10.1016/s0167-6393(00)00085-6

Iverson, P., & Krumhansl, C. (1993). Isolating the dynamic attributes of musical timbre. *J Acoust Soc Am*, *94*(5), 2595–2603. https://doi.org/10.1121/1.407371

Johnsrude, I., Penhune, V., & Zatorre, R. (2000). Functional specificity in the right human auditory cortex for perceiving pitch direction. *Brain*, *123*(1), 155–163. https://doi.org/10.1093/brain/123.1.155

Kaneshiro, S., Hiraumi, H., & Sato, H. (2020). Central processing of speech sounds and non-speech sounds with similar spectral distribution: An auditory evoked potential study. *Auris Nasus Larynx*, *47*(5), 727–733. https://doi.org/10.1016/j.anl.2020.02.008

Kavanagh, G., & Kelly, J. (1987). Contribution of auditory cortex to sound localization by the ferret (Mustela putorius). *J Neurophysiol*, *57*(6), 1746–1766. https://doi.org/10.1152/jn.1987.57.6.1746

Kazui, S. (1990). Subcortical auditory agnosia. *Brain Lang*, *38*(4), 476–487. https://doi.org/10.1016/0093-934x(90)90132-z

Kim, S., Overath, T., Sedley, W., Kumar, S., Teki, S., Kikuchi, Y., Patterson, R., & Griffiths, T. (2022). MEG correlates of temporal regularity relevant to pitch perception in human auditory cortex. *NeuroImage*, *249*, 118879. https://doi.org/10.1016/j.neuroimage.2022.118879

Krumbholz, K. (2003). Neuromagnetic Evidence for a Pitch Processing Center in Heschl's Gyrus. *Cereb Cortex*, *13*(7), 765–772. https://doi.org/10.1093/cercor/13.7.765

Krumbholz, K., Patterson, R., & Pressnitzer, D. (2000). The lower limit of pitch as determined by rate discrimination. *J Acoust Soc Am*, *108*(3), 1170–1180. https://doi.org/10.1121/1.1287843

Krumhansl, C. (1989). Why is musical timbre so hard to understand? In S. Nielzén & O. Olsson (Eds.), *Structure and Perception of Electroacoustic Sound and Music* (pp. 43–53, Vol. 846). Elsevier, Amsterdam. http://music.psych.cornell.edu/articles/timbre/Why_Is_Musical_Timbre_so%20hard_to_understand.pdf

Krumhansl, C., & Iverson, P. (1992). Perceptual interactions between musical pitch and timbre. *J Exp Psychol Hum Percept Perform*, *18*(3), 739–751. https://doi.org/10.1037/0096-1523.18.3.739

Kumar, S., Stephan, K., Warren, J., Friston, K., & Griffiths, T. (2007). Hierarchical Processing of Auditory Objects in Humans (P. Bourne, Ed.). *PLoS Comput Biol*, *3*(6), e100. https://doi.org/10.1371/journal.pcbi.0030100

Lartillot, O., Toiviainen, P., & Eerola, T. (2008). A Matlab Toolbox for Music Information Retrieval. In *Studies in classification, data analysis, and knowledge organization* (pp. 261–268). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-78246-9_31

Leonard, C., Puranik, C., Kuldau, J., & Lombardino, L. (1998). Normal variation in the frequency and location of human auditory cortex landmarks. heschl's gyrus: Where is it? *Cereb Cortex*, *8*(5), 397–406. https://doi.org/10.1093/cercor/8.5.397

Logan, B. (2000). Mel Frequency Cepstral Coefficients for Music Modeling. *International Society for Music Information Retrieval Conference*. https://api.semanticscholar.org/CorpusID:17454278

Loughran, R., Walker, J., O'Neill, M., & O'Farrell, M. (2008). Musical instrument identification using principalcomponent analysis and multi-layered perceptrons. *International Conference on Audio, Language and Image Processing (ICALIP)*, 643–648.

Lu, T., Liang, L., & Wang, X. (2001). Temporal and rate representations of time-varying signals in the auditory cortex of awake primates. *Nat Neurosci*, *4*(11), 1131–1138. https://doi.org/10.1038/nn737

Malhotra, S., & Lomber, S. (2007). Sound Localization During Homotopic and Heterotopic Bilateral Cooling Deactivation of Primary and Nonprimary Auditory Cortical Areas in the Cat. *J Neurophysiol*, *97*(1), 26–43. https://doi.org/10.1152/jn.00720.2006

Marozeau, J., & de Cheveigné, A. (2007). The effect of fundamental frequency on the brightness dimension of timbre. *J Acoust Soc Am*, *121*(1), 383–387. https://doi.org/10.1121/1.2384910

Marozeau, J., de Cheveigné, A., McAdams, S., & Winsberg, S. (2003). The dependency of timbre on fundamental frequency. *J Acoust Soc Am*, *114*(5), 2946–2957. https://doi.org/10.1121/1.1618239

McAdams, S. (1999). Perspectives on the Contribution of Timbre to Musical Structure. *Computer Music Journal*, *23*(3), 85–102. https://doi.org/10.1162/014892699559797

McAdams, S., Winsberg, S., Donnadieu, S., De Soete, G., & Krimphoff, J. (1995). Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. *Psychol Res*, *58*(3), 177–192. https://doi.org/10.1007/bf00419633

Meddis, R. (1988). Simulation of auditory–neural transduction: Further studies. *J Acoust Soc Am*, *83*(3), 1056–1063. https://doi.org/10.1121/1.396050

Melara, R., & Marks, L. (1990). Interaction among auditory dimensions: Timbre, pitch, and loudness. *Percept Psychophys*, *48*(2), 169–178. https://doi.org/10.3758/bf03207084

Moore, B. (2004). *An Introduction to the Psychology of Hearing* (5th ed.). Elsevier Academic Press, London.

Nelken, I., Bizley, J., Shamma, S., & Wang, X. (2014). Auditory Cortical Processing in Real-World Listening: The Auditory System Going Real. *J Neurosci*, *34*(46), 15135–15138. https://doi.org/10.1523/jneurosci.2989-14.2014

Obleser, J., Boecker, H., Drzezga, A., Haslinger, B., Hennenlotter, A., Roettinger, M., Eulitz, C., & Rauschecker, J. (2005). Vowel sound extraction in anterior superior temporal cortex. *Hum Brain Mapp*, *27*(7), 562–571. https://doi.org/10.1002/hbm.20201

Ogg, M., Slevc, L., & Idsardi, W. (2017). The time course of sound category identification: Insights from acoustic features. *J Acoust Soc Am*, *142*(6), 3459–3473. https://doi.org/10.1121/1.5014057

Ohl, F., & Scheich, H. (1997). Orderly cortical representation of vowels based onformantinteraction. *Proc Natl Acad Sci*, *94*(17), 9440–9444. https://doi.org/10.1073/pnas.94.17.9440

Patterson, R., Allerhand, M., & Giguère, C. (1995). Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform. *J Acoust Soc Am*, *98*(4), 1890–1894. https://doi.org/10.1121/1.414456

Patterson, R., Gaudrain, E., & Walters, T. (2010a). The Perception of Family and Register in Musical Tones. In *Springer handbook of auditory research* (pp. 13–50). Springer, New York. https://doi.org/10.1007/978-1-4419-6114-3_2

Patterson, R., & Holdsworth, J. (1991). A functional model of neural activity patterns and auditory images. In W. Ainsworth (Ed.), *Advancesin speech, hearing and language processing* (pp. 547–563, Vol. 3). JAI Press, London.

Patterson, R., & Irino, T. (2014). Size Matters in Hearing: How the Auditory System Normalizes the Sounds of Speech and Music for Source Size. In *Springer handbook of auditory research* (pp. 417–440). Springer, New York. https://doi.org/10.1007/978-1-4614-9102-6_23

Patterson, R., Robinson, K., Holdsworth, J., McKeown, D., Zhang, C., & Allerhand, M. (1992). Complex Sounds and Auditory Images. In *Auditory physiology and perception* (pp. 429–446). Elsevier. https://doi.org/10.1016/b978-0-08-041847-6.50054-x

Patterson, R., Smith, D., van Dinther, R., & Walters, T. (2008). Size Information in the Production and Perception of Communication Sounds. In W. Yost, A. Popper, & R. Fay (Eds.), *Springer handbook of auditory research* (pp. 43–75). Springer, New York. https://doi.org/10.1007/978-0-387-71305-2_3

Patterson, R., Uppenkamp, S., Johnsrude, I., & Griffiths, T. (2002). The Processing of Temporal Pitch and Melody Information in Auditory Cortex. *Neuron*, *36*(4), 767–776. https://doi.org/10.1016/s0896-6273(02)01060-7

Patterson, R., Walters, T., Monaghan, J., & Gaudrain, E. (2010b). Reviewing the Definition of Timbre as it Pertains to the Perception of Speech and Musical Sounds. In E. Lopez-Poveda, A. Palmer, & R. Meddis (Eds.), *The neurophysiological bases of auditory perception* (pp. 223–233). Springer, New York. https://doi.org/10.1007/978-1-4419-5686-6_21

Peeters, G. (2004). A large set of audio features for sound description (similarity and classification) in the CUIDADO project. *CUIDADO IST Project Report*, 1–25.

Penagos, H., Melcher, J., & Oxenham, . (2004). A Neural Representation of Pitch Salience in Nonprimary Human Auditory Cortex Revealed with Functional Magnetic Resonance Imaging. *J Neurosci*, *24*(30), 6810–6815. https://doi.org/10.1523/jneurosci.0383-04.2004

Pitt, M. (1994). Perception of pitch and timbre by musically trained and untrained listeners. *J Exp Psychol Hum Percept Perform*, *20*(5), 976–986. https://doi.org/10.1037/0096-1523.20.5.976

Plack, C., & Oxenham, A. (2005). Overview: The Present and Future of Pitch. In C. Plack, R. Fay, A. Oxenham, & A. Popper (Eds.), *Pitch. springer handbook of auditory research* (pp. 1–6, Vol. 24). Springer, New York. https://doi.org/10.1007/0-387-28958-5_1

Plomp, R. (1970). Timbre as a multidimensional attribute of complex tones. In R. Plomp & G. Smoorenburg (Eds.), *Frequency Analysis and Periodicity Detection in Hearing* (pp. 397–414). Sijthoff, Leiden.

Poeppel, D., & Hickok, G. (2015). Electromagnetic recording of the auditory system. In *Handbook of clinical neurology* (pp. 245–255). Elsevier. https://doi.org/10.1016/b978-0-444-62630-1.00014-7

Quiroga-Martinez, D., Hansen, N., Højlund, A., Pearce, M., Brattico, E., & Vuust, P. (2020). Decomposing neural responses to melodic surprise in musicians and non-musicians: Evidence for a hierarchy of predictions in the auditory system. *Neuroimage*, *215*, 116816. https://doi.org/10.1016/j.neuroimage.2020.116816

Regev, T., Markusfeld, G., Deouell, L., & Nelken, I. (2021). Context Sensitivity across Multiple Time scales with a Flexible Frequency Bandwidth. *Cereb Cortex*, *32*(1), 158–175. https://doi.org/10.1093/cercor/bhab200

Reuter, C. Out of Timbre Space - How to measure timbre similarity? In: Wien, 2019.

Reuter, C., Czedik-Eysenberg, I., Siddiq, S., & Oehler, M. (2017). Formanten als hilfreiche Timbre-Deskriptoren für die Darstellung von Blasinstrumentenklängen. *Fortschritte der Akustik–DAGA 2017, 43. Deutsche Jahrestagung für Akustik*, 190–193.

Reuter, C., Czedik-Eysenberg, I., Siddiq, S., & Oehler, M. (2018a). "... wenn das Gute liegt so nah": Instrumentale Formantnähe und Klangfarbenähnlichkeit aus menschlicher und rechnerischer Perspektive. *Fortschritte der Akustik–DAGA 2018, 44. Deutsche Jahrestagung für Akustik*, 1711–1714.

Reuter, C., Czedik-Eysenberg, I., Siddiq, S., & Oehler, M. (2018b). Formant Distances and the Similarity Perception of Wind Instrument Timbres. In R. Parncutt & S. Sattmann (Eds.), *Proceedings of ICMPC15/ESCOM10* (pp. 367–371).

Ritter, S., Dosch, H., Specht, H.-J., & Rupp, A. (2005). Neuromagnetic responses reflect the temporal pitch change of regular interval sounds. *Neuroimage*, *27*(3), 533–543. https://doi.org/10.1016/j.neuroimage.2005.05.003

Robin, D., Tranel, D., & Damasio, H. (1990). Auditory perception of temporal and spectral events in patients with focal left and right cerebral lesions. *Brain Lang*, *39*(4), 539–555. https://doi.org/10.1016/0093-934x(90)90161-9

Rupp, A., Spachmann, A., Dettlaff, A., & Patterson, R. (2013). Cortical Activity Associated with the Perception of Temporal Asymmetry in Ramped and Damped Noises. In *Basic aspects of hearing* (pp. 427–433). Springer, New York. https://doi.org/10.1007/978-1-4614-1590-9_47

Sakai, M., Chimoto, S., Qin, L., & Sato, Y. (2009). Differential representation of spectral and temporal information by primary auditory cortex neurons in awake cats: Relevance to auditory scene analysis. *Brain Res*, *1265*, 80–92. https://doi.org/10.1016/j.brainres.2009.01.064

Saldanha, E., & Corso, J. (1964). Timbre Cues and the Identification of Musical Instruments. *J Acoust Soc Am*, *36*(11), 2021–2026. https://doi.org/10.1121/1.1919317

Samson, S., & Zatorre, R. (1988). Melodic and harmonic discrimination following unilateral cerebral excision. *Brain Cogn*, *7*(3), 348–360. https://doi.org/10.1016/0278-2626(88)90008-5

Samson, S., & Zatorre, R. (1994). Contribution of the right temporal lobe to musical timbre discrimination. *Neuropsychologia*, *32*(2), 231–240. https://doi.org/10.1016/0028-3932(94)90008-6

Samson, S., Zatorre, R., & Ramsay, J. (2002). Deficits of musical timbre perception after unilateral temporal-lobe lesion revealed with multidimensional scaling. *Brain*, *125*(3), 511–523. https://doi.org/10.1093/brain/awf051

Scharinger, M., Idsardi, W., & Poe, S. (2011). A Comprehensive Three-dimensional Cortical Map of Vowel Space. *J Cogn Neurosci*, *23*(12), 3972–3982. https://doi.org/10.1162/jocn_a_00056

Schneider, P., Sluming, V., Roberts, N., Scherg, M., Goebel, R., Specht, H., Dosch, H., Bleeck, S., Stippich, C., & Rupp, A. (2005). Structural and functional asymmetry of lateral Heschl's gyrus reflects pitch perception preference. *Nat Neurosci*, *8*(9), 1241–1247. https://doi.org/10.1038/nn1530

Schouten, J. (1970). Timbre as a multidimensional attribute of complex tones. In R. Plomp & G. Smoorenburg (Eds.), *Frequency analysis and periodicity detection in hearing* (pp. 397–414). Sijthoff, Leiden.

Schumann, K. (1929). *Physik der Klangfarben* (Vol. 2). Breitkopf; Haertel, Leipzig.

Shamma, S., Fleshman, J., Wiser, P., & Versnel, H. (1993). Organization of response areas in ferret primary auditory cortex. *J Neurophysiol*, *69*(2), 367–383. https://doi.org/10.1152/jn.1993.69.2.367

Siddiq, S., & Reuter, C. Klangfarbe in 3D – Lost in Timbre Space. In: In *"Musik und Familie" - Jahrestagung der Deutschen Gesellschaft für Musikpsychologie (DGM)*. Frankfurt, 2013.

Siddiq, S., Reuter, C., & Czedik-Eysenberg, I. (2014). Kein Raum für Klangfarben – Timbre Spaces im Vergleich. *Fortschritte der Akustik–DAGA 2014, 40. Deutsche Jahrestagung für Akustik*, 56–57.

Siddiq, S., Reuter, C., Czedik-Eysenberg, I., & Knauf, D. (2015). Vergleichende Untersuchungen zu Timbre Space Studien. *Fortschritte der Akustik–DAGA 2015, 41. Deutsche Jahrestagung für Akustik*, 811–813.

Siddiq, S., Reuter, C., Czedik-Eysenberg, I., & Knauf, D. (2017). Timbre Space reloaded - Tonhöhe und Dynamik als Teil der Klangfarbenempfindung. *Fortschritte der Akustik–DAGA 2017, 43. Deutsche Jahrestagung für Akustik*, 194–197.

Siddiq, S., Reuter, C., Czedik-Eysenberg, I., & Knauf, D. (2018). Die physikalischen Korrelate von Instrumentalklangfarben. *Fortschritte der Akustik–DAGA 2018, 44. Deutsche Jahrestagung für Akustik*, 1695–1698.

Sidtis, J., & Volpe, B. (1988). Selective loss of complex-pitch or speech discrimination after unilateral lesion. *Brain Lang*, *34*(2), 235–245. https://doi.org/10.1016/0093-934x(88)90135-6

Siedenburg, K. (2019). Specifying the perceptual relevance of onset transients for musical instrument identification. *J Acoust Soc Am*, *145*(2), 1078–1087. https://doi.org/10.1121/1.5091778

Siedenburg, K., & McAdams, S. (2017). Four Distinctions for the Auditory "Wastebasket" of Timbre. *Front Psychol*, *8*. https://doi.org/10.3389/fpsyg.2017.01747

Smith, A., Parsons, C., Lanyon, R., Bizley, J., Akerman, C., Baker, G., Dempster, A., Thompson, I., & King, A. (2004). An investigation of the role of auditory cortex in sound localization using muscimol-releasing Elvax. *Eur J Neurosci*, *19*(11), 3059–3072. https://doi.org/10.1111/j.0953-816x.2004.03379.x

Smith, D., Patterson, R., Turner, R., Kawahara, H., & Irino, T. (2005). The processing and perception of size information in speech sounds. *J Acoust Soc Am*, *117*(1), 305–318. https://doi.org/10.1121/1.1828637

Staeren, N., Renvall, H., De Martino, F., Goebel, R., & Formisano, E. (2009). Sound Categories Are Represented as Distributed Patterns in the Human Auditory Cortex. *Curr Biol*, *19*(6), 498–502. https://doi.org/10.1016/j.cub.2009.01.066

Stewart, L., von Kriegstein, K., Warren, J., & Griffiths, T. (2006). Music and the brain: Disorders of musical listening. *Brain*, *129*(10), 2533–2553. https://doi.org/10.1093/brain/awl171

Stumpf, C. (1890). *Tonpsychologie* (Vol. 2). Hirzel, Stuttgart.

Suied, C., Agus, T., Thorpe, S., Mesgarani, N., & Pressnitzer, D. (2014). Auditory gist: Recognition of very short sounds from timbre cues. *J Acoust Soc Am*, *135*(3), 1380–1391. https://doi.org/10.1121/1.4863659

Town, S., & Bizley, J. (2013). Neural and behavioral investigations into timbre perception. *Front Syst Neurosci*, *7*. https://doi.org/10.3389/fnsys.2013.00088

Tramo, M., Shah, G., & Braida, L. (2002). Functional Role of Auditory Cortex in Frequency Processing and Pitch Perception. *J Neurophysiol*, *87*(1), 122–139. https://doi.org/10.1152/jn.00104.1999

van Dinther, R., & Patterson, R. (2006). Perception of acoustic scale and size in musical instrument sounds. *J Acoust Soc Am*, *120*(4), 2158–2176. https://doi.org/10.1121/1.2338295

von Helmholtz, H. (1863). *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik*. F. Vieweg, Braunschweig.

von Helmholtz, H. (1895). *On the Sensations of Tone as a Physiological Basis for the Theory of Music* (A. Translated by Ellis, Ed.; 3rd ed.). Longmans, Green, Co., London, New York. https://archive.org/details/onsensationsofto00helmrich/page/n8

von Kriegstein, K., Smith, D., Patterson, R., Ives, D., & Griffiths, T. (2007). Neural Representation of Auditory Size in the Human Voice and in Sounds from Other Resonant Sources. *Curr Biol*, *17*(13), 1123–1128. https://doi.org/10.1016/j.cub.2007.05.061

von Kriegstein, K., Smith, D., Patterson, R., Kiebel, S., & Griffiths, T. (2010). How the Human Brain Recognizes Speech in the Context of Changing Speakers. *J Neurosci*, *30*(2), 629–638. https://doi.org/10.1523/jneurosci.2742-09.2010

von Kriegstein, K., Warren, J., Ives, D., Patterson, R., & Griffiths, T. (2006). Processing the acoustic effect of size in speech sounds. *Neuroimage*, *32*(1), 368–375. https://doi.org/10.1016/j.neuroimage.2006.02.045

Wallace, M., & Palmer, A. (2009). Functional subdivisions in low-frequency primary auditory cortex (AI). *Exp Brain Res*, *194*(3), 395–408. https://doi.org/10.1007/s00221-009-1714-8

Warren, J., Jennings, A., & Griffiths, T. (2005). Analysis of the spectral envelope of sounds by the human brain. *Neuroimage*, *24*(4), 1052–1057. https://doi.org/10.1016/j.neuroimage.2004.10.031

Warren, J., Uppenkamp, S., Patterson, R., & Griffiths, T. (2003). Analyzing Pitch Chroma and Pitch Height in the Human Brain. *Ann N Y Acad Sci*, *999*(1), 212–214. https://doi.org/10.1196/annals.1284.032

Warrier, C., & Zatorre, R. (2004). Right temporal cortex is critical for utilization of melodic contextual cues in a pitch constancy task. *Brain*, *127*(7), 1616–1625. https://doi.org/10.1093/brain/awh183

Wedin, L., & Goude, G. (1972). Dimension analysis of the perception of instrumental timbre. *Scand J Psychol*, *13*(1), 228–240. https://doi.org/10.1111/j.1467-9450.1972.tb00071.x

Wei, Y., Gan, L., & Huang, X. (2022). A Review of Research on the Neurocognition for Timbre Perception. *Front Psychol*, *13*. https://doi.org/10.3389/fpsyg.2022.869475

Wessel, D. (1979). Timbre Space as a Musical Control Structure. *Comput Music J*, *3*(2), 45. https://doi.org/10.2307/3680283

Whitfield, I. (1980). Auditory cortex and the pitch of complex tones. *J Acoust Soc Am*, *67*(2), 644–647. https://doi.org/10.1121/1.383889

Yost, W. (1996). Pitch of iterated rippled noise. *J Acoust Soc Am*, *100*(1), 511–518. https://doi.org/10.1121/1.415873

Zatorre, R. (1988). Pitch perception of complex tones and human temporal-lobe function. *J Acoust Soc Am*, *84*(2), 566–572. https://doi.org/10.1121/1.396834

Zatorre, R., Evans, A., & Meyer, E. (1994). Neural mechanisms underlying melodic perception and memory for pitch. *J Neurosci*, *14*(4), 1908–1919. https://doi.org/10.1523/jneurosci.14-04-01908.1994

Zatorre, R., & Penhune, V. (2001). Spatial Localization after Excision of Human Auditory Cortex. *J Neurosci*, *21*(16), 6321–6328. https://doi.org/10.1523/jneurosci.21-16-06321.2001

Zatorre, R., & Salimpoor, V. (2013). From perception to pleasure: Music and its neural substrates. *Proc Natl Acad Sci*, *110*(supplement$_2$), 10430–10437. https://doi.org/10.1073/pnas.1301228110

8 Personal publications

Publications in the context of this work

Günther, M., Andermann, M., Reuter, C., & Rupp, A. (2021). Neuromagnetische Repräsentation von Timbre im auditorischen Cortex. *Fortschritte der Akustik–DAGA 2021, 47. Deutsche Jahrestagung für Akustik*, 528–530.

This conference paper presents the first pilot data of the current study and highlights the commonalities and differences between this study and a previous study by Dr. Martin Andermann (2017) who investigated the processing of voice pitch, vowel type and speaker size in the human auditory cortex.

Other publications

Andermann, M., Günther, M., Patterson, R., & Rupp, A. (2021). Early cortical processing of pitch height and the role of adaptation and musicality. *Neuroimage, 225*, 117501. https://doi.org/10.1016/j.neuroimage.2020.117501

Appendix

Table 25: Melodies with contour change that were used as stimuli in the MEG experiment. Of the 16 melodies with contour change, 8 melodies had natural attack times (natural) and 8 melodies had unified attack times (unified). Of the 8 melodies with natural attack times and the 8 melodies with unified attack times, each 2 melodies were played by the same instrument(s) but differed in sound location. Instruments are given for the two tone triplets (tones 1–3, tones 3–6) that made up the respective melody. Sound location (*here:* Location) is given as R (right side) and L (left side) for both tone triplets, i.e. RR means that the sound of both tone triplets was on the right side. For a better discrimination of the melodies applied in MEG run A (natural attack times) and in MEG run B (unified attack times), melodies with natural attack times are marked with *a* and melodies with unified attack times are marked with *b*.

Melody		**Instrument**			
Name	**Change**	**Tones 1–3**	**Tones 4–6**	**Attack**	**Location**
OOr^{a}	contour	Oboe	Oboe	natural	RR
OOl^{a}	contour	Oboe	Oboe	natural	LL
OOr^{b}	contour	Oboe	Oboe	unified	RR
OOl^{b}	contour	Oboe	Oboe	unified	LL
BBr^{a}	contour	Bassoon	Bassoon	natural	RR
BBl^{a}	contour	Bassoon	Bassoon	natural	LL
BBr^{b}	contour	Bassoon	Bassoon	unified	RR
BBl^{b}	contour	Bassoon	Bassoon	unified	LL
CCr^{a}	contour	Clarinet	Clarinet	natural	RR
CCl^{a}	contour	Clarinet	Clarinet	natural	LL
CCr^{b}	contour	Clarinet	Clarinet	unified	RR
CCl^{b}	contour	Clarinet	Clarinet	unified	LL
KKr^{a}	contour	Bass Clarinet	Bass Clarinet	natural	RR
KKl^{a}	contour	Bass Clarinet	Bass Clarinet	natural	LL
KKr^{b}	contour	Bass Clarinet	Bass Clarinet	unified	RR
KKl^{b}	contour	Bass Clarinet	Bass Clarinet	unified	LL

Table 26: Melodies with register change that were used as stimuli in the MEG experiment. Of the 16 melodies with register change, 8 melodies had natural attack times (natural) and 8 melodies had unified attack times (unified). Of the 8 melodies with natural attack times and the 8 melodies with unified attack times, each 2 melodies were played by the same instrument(s) but differed in sound location. Instruments are given for the two tone triplets (tones 1–3, tones 3–6) that made up the respective melody. Sound location (*here:* Location) is given as R (right side) and L (left side) for both tone triplets, i.e. RR means that the sound of both tone triplets was on the right side. For a better discrimination of the melodies applied in MEG run A (natural attack times) and in MEG run B (unified attack times), melodies with natural attack times are marked with *a* and melodies with unified attack times are marked with *b*.

Melody		**Instrument**			
Name	**Change**	**Tones 1–3**	**Tones 4–6**	**Attack**	**Location**
OBr^{a}	register	Oboe	Bassoon	natural	RR
OBl^{a}	register	Oboe	Bassoon	natural	LL
OBr^{b}	register	Oboe	Bassoon	unified	RR
OBl^{b}	register	Oboe	Bassoon	unified	LL
BOr^{a}	register	Bassoon	Oboe	natural	RR
BOl^{a}	register	Bassoon	Oboe	natural	LL
BOr^{b}	register	Bassoon	Oboe	unified	RR
BOl^{b}	register	Bassoon	Oboe	unified	LL
CKr^{a}	register	Clarinet	Bass Clarinet	natural	RR
CKl^{a}	register	Clarinet	Bass Clarinet	natural	LL
CKr^{b}	register	Clarinet	Bass Clarinet	unified	RR
CKl^{b}	register	Clarinet	Bass Clarinet	unified	LL
KCr^{a}	register	Bass Clarinet	Clarinet	natural	RR
KCl^{a}	register	Bass Clarinet	Clarinet	natural	LL
KCr^{b}	register	Bass Clarinet	Clarinet	unified	RR
KCl^{b}	register	Bass Clarinet	Clarinet	unified	LL

Table 27: Melodies with family change that were used as stimuli in the MEG experiment. Of the 16 melodies with family change, 8 melodies had natural attack times (natural) and 8 melodies had unified attack times (unified). Of the 8 melodies with natural attack times and the 8 melodies with unified attack times, each 2 melodies were played by the same instrument(s) but differed in sound location. Instruments are given for the two tone triplets (tones 1–3, tones 3–6) that made up the respective melody. Sound location (*here:* Location) is given as R (right side) and L (left side) for both tone triplets, i.e. RR means that the sound of both tone triplets was on the right side. For a better discrimination of the melodies applied in MEG run A (natural attack times) and in MEG run B (unified attack times), melodies with natural attack times are marked with *a* and melodies with unified attack times are marked with *b*.

Melody		**Instrument**			
Name	**Change**	**Tones 1–3**	**Tones 4–6**	**Attack**	**Location**
OCr^{a}	family	Oboe	Clarinet	natural	RR
OCl^{a}	family	Oboe	Clarinet	natural	LL
OCr^{b}	family	Oboe	Clarinet	unified	RR
OCl^{b}	family	Oboe	Clarinet	unified	LL
COr^{a}	family	Clarinet	Oboe	natural	RR
COl^{a}	family	Clarinet	Oboe	natural	LL
COr^{b}	family	Clarinet	Oboe	unified	RR
COl^{b}	family	Clarinet	Oboe	unified	LL
BKr^{a}	family	Bassoon	Bass Clarinet	natural	RR
BKl^{a}	family	Bassoon	Bass Clarinet	natural	LL
BKr^{b}	family	Bassoon	Bass Clarinet	unified	RR
BKl^{b}	family	Bassoon	Bass Clarinet	unified	LL
KBr^{a}	family	Bass Clarinet	Bassoon	natural	RR
KBl^{a}	family	Bass Clarinet	Bassoon	natural	LL
KBr^{b}	family	Bass Clarinet	Bassoon	unified	RR
KBl^{b}	family	Bass Clarinet	Bassoon	unified	LL

Table 28: Melodies with sound location change that were used as stimuli in the MEG experiment. Of the 16 melodies with sound location change, 8 melodies had natural attack times (natural) and 8 melodies had unified attack times (unified). Of the 8 melodies with natural attack times and the 8 melodies with unified attack times, each 2 melodies were played by the same instrument(s) but differed in sound location. Instruments are given for the two tone triplets (tones 1–3, tones 3–6) that made up the respective melody. Sound location (*here:* Location) is given as R (right side) and L (left side) for both tone triplets, i.e. RL means that the sound of the first triplet was on the right side, whereas the sound of the second triplet was on the left side. For a better discrimination of the melodies applied in MEG run A (natural attack times) and in MEG run B (unified attack times), melodies with natural attack times are marked with *a* and melodies with unified attack times are marked with *b*.

Melody		**Instrument**			
Name	**Change**	**Tones 1–3**	**Tones 4–6**	**Attack**	**Location**
OOrl^{a}	location	Oboe	Oboe	natural	RL
OOlr^{a}	location	Oboe	Oboe	natural	LR
OOrl^{b}	location	Oboe	Oboe	unified	RL
OOlr^{b}	location	Oboe	Oboe	unified	LR
BBrl^{a}	location	Bassoon	Bassoon	natural	RL
BBlr^{a}	location	Bassoon	Bassoon	natural	LR
BBrl^{b}	location	Bassoon	Bassoon	unified	RL
BBlr^{b}	location	Bassoon	Bassoon	unified	LR
CCrl^{a}	location	Clarinet	Clarinet	natural	RL
CClr^{a}	location	Clarinet	Clarinet	natural	LR
CCrl^{b}	location	Clarinet	Clarinet	unified	RL
CClr^{b}	location	Clarinet	Clarinet	unified	LR
KKrl^{a}	location	Bass Clarinet	Bass Clarinet	natural	RL
KKlr^{a}	location	Bass Clarinet	Bass Clarinet	natural	LR
KKrl^{b}	location	Bass Clarinet	Bass Clarinet	unified	RL
KKlr^{b}	location	Bass Clarinet	Bass Clarinet	unified	LR

Table 29: Individual experimental conditions that were applied for the data analysis of the two MEG runs A and B. For a better discrimination of the melodies applied in MEG run A (natural attack times) and in MEG run B (unified attack times), melodies with natural attack times are marked with *a* and melodies with unified attack times are marked with *b*. For each MEG run, the 16 individual conditions comprised four *cont* conditions summarising the respective melodies with contour change, four *reg* conditions containing the respective melodies with register change, four *fam* conditions summarising the respective melodies with family change, and four *loc* conditions containing the respective melodies with sound location change.

MEG run a			**MEG run b**		
Condition	**Melodies**		**Condition**	**Melodies**	
cont_OO^{a}	**OO**r^{a}	**OO**l^{a}	cont_OO^{b}	**OO**r^{b}	**OO**l^{b}
cont_BB^{a}	**BB**r^{a}	**BB**l^{a}	cont_BB^{b}	**BB**r^{b}	**BB**l^{b}
cont_CC^{a}	**CC**r^{a}	**CC**l^{a}	cont_CC^{b}	**CC**r^{b}	**CC**l^{b}
cont_KK^{a}	**KK**r^{a}	**KK**l^{a}	cont_BB^{b}	**KK**r^{b}	**KK**l^{b}
reg_OB^{a}	**OB**r^{a}	**OB**l^{a}	reg_OB^{b}	**OB**r^{b}	**OB**l^{b}
reg_BO^{a}	**BO**r^{a}	**BO**l^{a}	reg_BO^{b}	**BO**r^{b}	**BO**l^{b}
reg_CK^{a}	**CK**r^{a}	**CK**l^{a}	reg_CK^{b}	**CK**r^{b}	**CK**l^{b}
reg_KC^{a}	**KC**r^{a}	**KC**l^{a}	reg_KC^{b}	**KC**r^{b}	**KC**l^{b}
fam_OC^{a}	**OC**r^{a}	**OC**l^{a}	fam_OC^{b}	**OC**r^{b}	**OC**l^{b}
fam_CO^{a}	**CO**r^{a}	**CO**l^{a}	fam_CO^{b}	**CO**r^{b}	**CO**l^{b}
fam_BK^{a}	**BK**r^{a}	**BK**l^{a}	fam_BK^{b}	**BK**r^{b}	**BK**l^{b}
fam_KB^{a}	**KB**r^{a}	**KB**l^{a}	fam_KB^{b}	**KB**r^{b}	**KB**l^{b}
loc_OO^{a}	**OO**rl^{a}	**OO**lr^{a}	loc_OO^{b}	**OO**rl^{b}	**OO**lr^{b}
loc_BB^{a}	**BB**rl^{a}	**BB**lr^{a}	loc_BB^{b}	**BB**rl^{b}	**BB**lr^{b}
loc_CC^{a}	**CC**rl^{a}	**CC**lr^{a}	loc_CC^{b}	**CC**rl^{b}	**CC**lr^{b}
loc_KK^{a}	**KK**rl^{a}	**KK**lr^{a}	loc_KK^{b}	**KK**rl^{b}	**KK**lr^{b}

Table 30: Pooled experimental conditions that were applied for the data analysis of the two MEG runs A and B. For each MEG run, the five pooled conditions comprised one *cont* condition summarising all melodies with contour change, one *reg* condition containing all melodies with register change, one *fam* condition summarising all melodies with family change, one *loc* condition containing all melodies with sound location change, and one *attack* condition summarising all melodies, irrespective of the change type. This table lists the 10 pooled conditions as they were used for statistical data analysis, i.e. the conditions $cont^a$, reg^a, fam^a, loc^a of MEG run A and the conditions $cont^b$, reg^b, fam^b, loc^b of MEG run B were summarised to *cont*, *reg*, *fam* and *loc*. The same applies to the respective individual conditions. Regarding the *all* condition, the differentiation between MEG run A and MEG run B was maintained to compare the neuromagnetic responses between the two MEG runs.

Condition	**Individual conditions**			
cont	cont_OO	cont_BB	cont_CC	cont_KK
reg	reg_OB	reg_BO	reg_CK	reg_KC
fam	fam_OC	fam_CO	fam_BK	fam_KB
loc	loc_OO	loc_BB	loc_CC	loc_KK
all	$cont^a$	reg^a	fam^a	loc^a
all	$cont^b$	reg^b	fam^b	loc^b

Table 31: Overview of the performed statistical analyses (main effects). For each analysis, the variable, the respective experimental conditions and dipole models are listed. The statistical tests were performed to analyse differences between the change types contour, register, family and sound location (*here:* sound loc), between natural and unified attack time, between left and right hemisphere, between low AMMA listeners (*lowA*) and high AMMA listeners (*highA*), and between fundamental listeners (*fund*) and overtone listeners (*over*).

Variable	**Conditions**		**Models**	
contour	**cont**	$\neq$ **cont**	CONT	REG FAM LOC
register	**reg**	$\neq$ **reg**	REG	CONT FAM LOC
family	**fam**	$\neq$ **fam**	FAM	CONT REG LOC
sound loc	**loc**	$\neq$ **loc**	LOC	CONT REG FAM
attack time	**all**	**all**	ATTACK^a	ATTACK^b
hemisphere	left	right	CONT REG FAM LOC	CONT REG FAM LOC
AMMA	**lowA**	**highA**	CONT REG FAM LOC	CONT REG FAM LOC
PITCHT	fund	**over**	CONT REG FAM LOC	CONT REG FAM LOC

Table 32: Overview of the performed statistical analyses (interaction effects). For each analysis, the variables, the respective experimental conditions and dipole models are listed. The statistical tests were performed to analyse interaction effects between the change types contour, register, family, sound location (*here:* sound loc) and the AMMA score, and between the same four change types and the PITCHT score.

Variables		**Conditions**				**Models**	
contour	x AMMA	**cont**	x AMMA	≠ **cont**	x AMMA	CONT	REG FAM LOC
register	x AMMA	**reg**	x AMMA	≠ **reg**	x AMMA	REG	CONT FAM LOC
family	x AMMA	**fam**	x AMMA	≠ **fam**	x AMMA	FAM	CONT REG LOC
sound loc	x AMMA	**loc**	x AMMA	≠ **loc**	x AMMA	LOC	CONT REG FAM
contour	x PITCHT	**cont**	x PITCHT	≠ **cont**	x PITCHT	CONT	REG FAM LOC
register	x PITCHT	**reg**	x PITCHT	≠ **reg**	x PITCHT	REG	CONT FAM LOC
family	x PITCHT	**fam**	x PITCHT	≠ **fam**	x PITCHT	FAM	CONT REG LOC
sound loc	x PITCHT	**loc**	xPITCHT	≠ **loc**	x PITCHT	LOC	CONT REG FAM

Table 33: Tone pairs that were presented once in the psychoacoustic test. Overall, there were 16 groups of tone pairs: four groups of tone pairs with pitch change (Pitch), four groups of tone pairs with pitch & register change (pRegister), four groups of tone pairs with pitch & family change (pFamily), and four groups of tone pairs with pitch & register & family change (pRegFam). Each group comprised six tone pairs that were played by one or two of four instruments (O = oboe, B = bassoon, C = clarinet, K = bass clarinet). The tones played by the oboe and the clarinet were D4, F♯4, A4; the tones played by the bassoon and the bass clarinet were D2, F♯2, A2. For simplicity, the register numbers were omitted and the tones are only given as D, F♯ and A.

	Instrument							
Condition	**Tone 1**	**Tone 2**	**Tone pairs**					
Pitch_OO	Oboe	Oboe	DF♯	F♯D	DA	AD	F♯A	AF♯
Pitch_BB	Bassoon	Basson	DF♯	F♯D	DA	AD	F♯A	AF♯
Pitch_CC	Clarinet	Clarinet	DF♯	F♯D	DA	AD	F♯A	AF♯
Pitch_KK	BClarinet	BClarinet	DF♯	F♯D	DA	AD	F♯A	AF♯
pRegister_OB	Oboe	Bassoon	DF♯	F♯D	DA	AD	F♯A	AF♯
pRegister_BO	Bassoon	Oboe	DF♯	F♯D	DA	AD	F♯A	AF♯
pRegister_CK	Clarinet	BClarinet	DF♯	F♯D	DA	AD	F♯A	AF♯
pRegister_KC	BClarinet	Clarinet	DF♯	F♯D	DA	AD	F♯A	AF♯
pFamily_OC	Oboe	Clarinet	DF♯	F♯D	DA	AD	F♯A	AF♯
pFamily_CO	Clarinet	Oboe	DF♯	F♯D	DA	AD	F♯A	AF♯
pFamily_BK	Bassoon	BClarinet	DF♯	F♯D	DA	AD	F♯A	AF♯
pFamily_KB	BClarinet	Bassoon	DF♯	F♯D	DA	AD	F♯A	AF♯
pRegFam_OK	Oboe	BClarinet	DF♯	F♯D	DA	AD	F♯A	AF♯
pRegFam_KO	BClarinet	Oboe	DF♯	F♯D	DA	AD	F♯A	AF♯
pRegFam_CB	Clarinet	Bassoon	DF♯	F♯D	DA	AD	F♯A	AF♯
pRegFam_BC	Bassoon	Clarinet	DF♯	F♯D	DA	AD	F♯A	AF♯

Table 34: Tone pairs that were presented twice in the psychoacoustic test. In total, there were 12 groups of tone pairs: four groups of tone pairs with register change (Register), four groups of tone pairs with family change (Family), and four groups of tone pairs with register & family change (RegFam). Each group comprised three tone pairs that were played by one or two of four instruments (O = oboe, B = bassoon, C = clarinet, K = bass clarinet). The tones played by the oboe and the clarinet were D4, F♯4, A4; the tones played by the bassoon and the bass clarinet were D2, F♯2, A2. For simplicity, the register numbers were omitted and the tones are only given as D, F♯ and A.

	Instrument				
Condition	**Tone 1**	**Tone 2**	**Tone pairs**		
Register_OB	Oboe	Bassoon	DD	F♯F♯	AA
Register_BO	Bassoon	Oboe	DD	F♯F♯	AA
Register_CK	Clarinet	BClarinet	DD	F♯F♯	AA
Register_KC	BClarinet	Clarinet	DD	F♯F♯	AA
Family_OC	Oboe	Clarinet	DD	F♯F♯	AA
Family_CO	Clarinet	Oboe	DD	F♯F♯	AA
Family_BK	Bassoon	BClarinet	DD	F♯F♯	AA
Family_KB	BClarinet	Bassoon	DD	F♯F♯	AA
RegFam_OK	Oboe	BClarinet	DD	F♯F♯	AA
RegFam_KO	BClarinet	Oboe	DD	F♯F♯	AA
RegFam_CB	Clarinet	Bassoon	DD	F♯F♯	AA
RegFam_BC	Bassoon	Clarinet	DD	F♯F♯	AA

Table 35: Further information on the NEURO and PSYCH subject groups.

	NEURO	**PSYCH**
Training intensity [hours per week]		
<10 years	2.2 ± 2.1	2.4 ± 2.1
10–14 years	3.6 ± 3.4	3.4 ± 4.0
15–18 years	6.0 ± 8.5	3.5 ± 3.9
last 6 months	1.5 ± 2.6	1.0 ± 1.7
Number of instruments		
$= 0$	7	6
$= 1$	12	6
$= 2$	5	1
≥ 3	8	4
Main instruments		
wind	3	2
plucked string	7	3
bowed string	5	2
percussion	2	2
keyboard	8	2

Table 36: Bootstrap results on the comparison between the $N1_{cng}/P2_{cng}$ responses to the fourth tone of the melodies (*tone 4*) and the other N1 and P2 responses to the second (*tone 2*), third (*tone 3*), fifth (*tone 5*), and sixth (*tone 6*) tones of the melody.

		N1 amplitude		**P2 amplitude**	
Comparison		**Effect size**	***p***	**Effect size**	***p***
tone 4	tone 2	0.1700	0.3040 n.s.	0.6151	0.0460*
tone 4	tone 3	0.4742	0.0405*	0.5940	0.0385*
tone 4	tone 5	-0.0121	0.5375 n.s.	0.3850	0.0405*
tone 4	tone 6	0.2380	0.2070 n.s.	0.8796	0.0020**

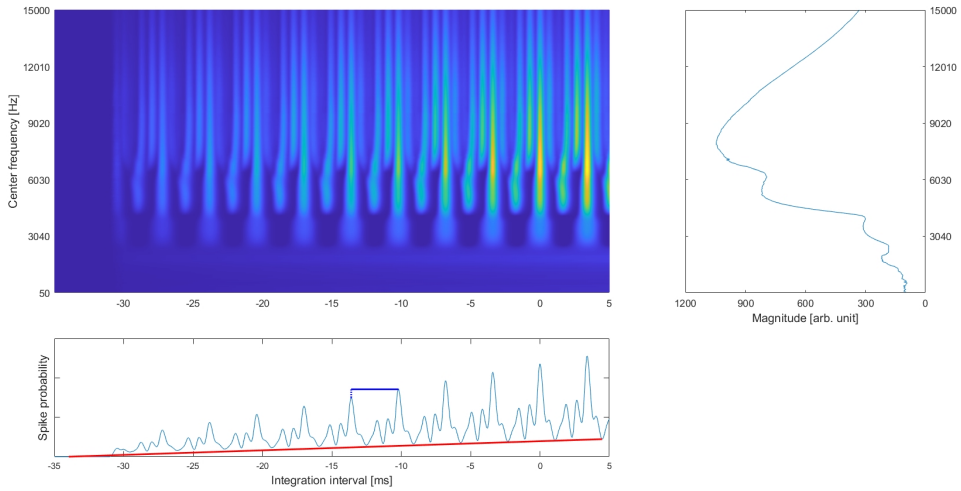


Figure 41: Stabilised auditory image of the oboe tone D4. The upper left subfigure represents the neural activity in 500 basilar membrane regions or frequency channels (characterised by center frequency in [Hz]) per time in [ms]. Dark colours indicate low neural activity, bright colours indicate high neural activity. The upper right subfigure depicts the magnitude in arbitrary units per frequency channel. The lower subfigure shows the spiking probability across the temporal integration interval in [ms]. Note that the first 5 ms (i.e. from -35 to -30 ms) represent a baseline interval and the remaining 35 ms (i.e. from -30 to 5 ms) show the relevant data. The extracted sound features periodicity and hissiness are represented by the blue and red lines, respectively.

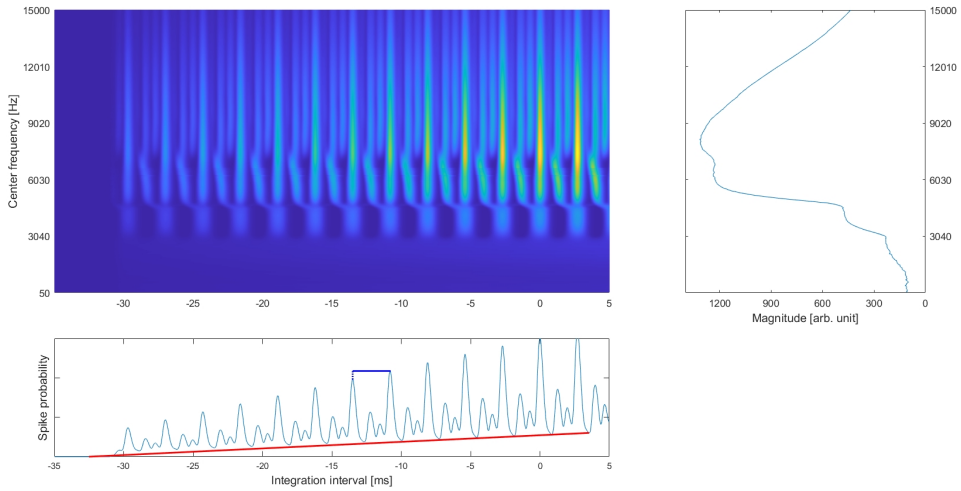


Figure 42: Stabilised auditory image of the oboe tone F♯4. The upper left subfigure represents the neural activity in 500 basilar membrane regions or frequency channels (characterised by center frequency in [Hz]) per time in [ms]. Dark colours indicate low neural activity, bright colours indicate high neural activity. The upper right subfigure depicts the magnitude in arbitrary units per frequency channel. The lower subfigure shows the spiking probability across the temporal integration interval in [ms]. Note that the first 5 ms (i.e. from -35 to -30 ms) represent a baseline interval and the remaining 35 ms (i.e. from -30 to 5 ms) show the relevant data. The extracted sound features periodicity and hissiness are represented by the blue and red lines, respectively.

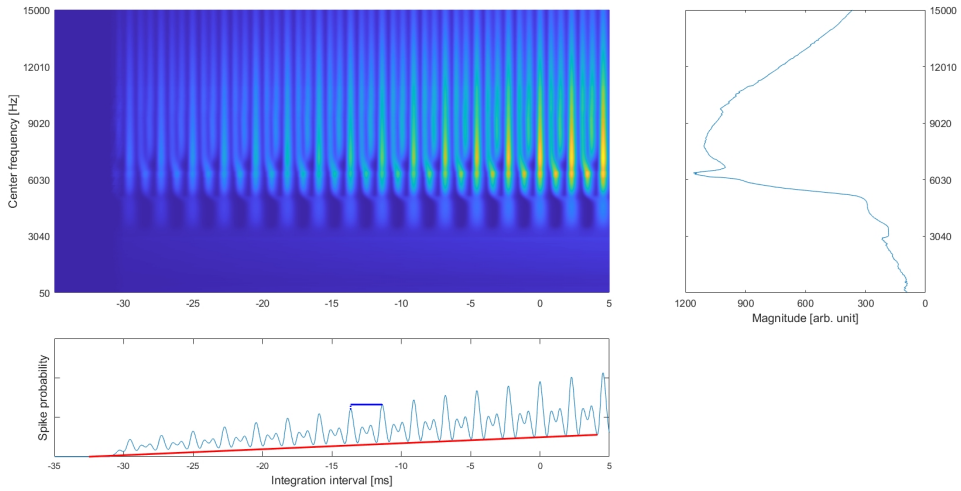


Figure 43: Stabilised auditory image of the oboe tone A4. The upper left subfigure represents the neural activity in 500 basilar membrane regions or frequency channels (characterised by center frequency in [Hz]) per time in [ms]. Dark colours indicate low neural activity, bright colours indicate high neural activity. The upper right subfigure depicts the magnitude in arbitrary units per frequency channel. The lower subfigure shows the spiking probability across the temporal integration interval in [ms]. Note that the first 5 ms (i.e. from -35 to -30 ms) represent a baseline interval and the remaining 35 ms (i.e. from -30 to 5 ms) show the relevant data. The extracted sound features periodicity and hissiness are represented by the blue and red lines, respectively.

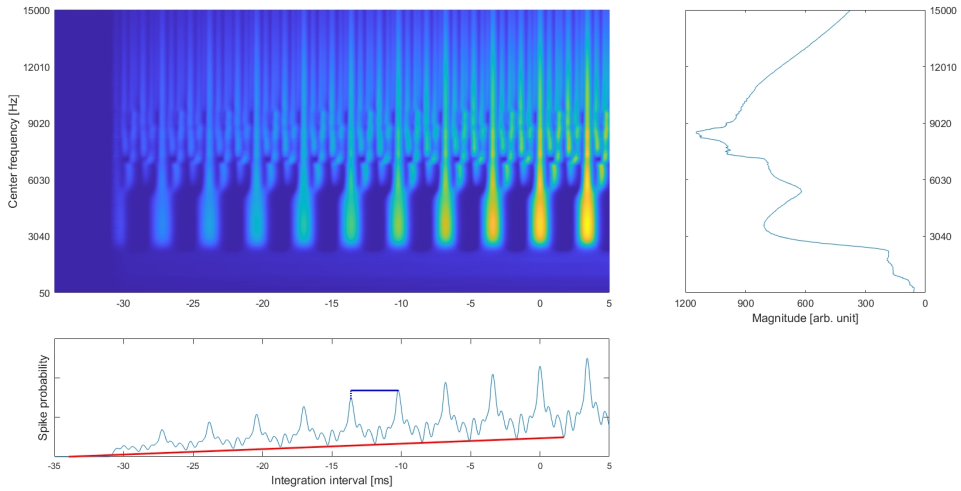


Figure 44: Stabilised auditory image of the clarinet tone D4. The upper left subfigure represents the neural activity in 500 basilar membrane regions or frequency channels (characterised by center frequency in [Hz]) per time in [ms]. Dark colours indicate low neural activity, bright colours indicate high neural activity. The upper right subfigure depicts the magnitude in arbitrary units per frequency channel. The lower subfigure shows the spiking probability across the temporal integration interval in [ms]. Note that the first 5 ms (i.e. from -35 to -30 ms) represent a baseline interval and the remaining 35 ms (i.e. from -30 to 5 ms) show the relevant data. The extracted sound features periodicity and hissiness are represented by the blue and red lines, respectively.

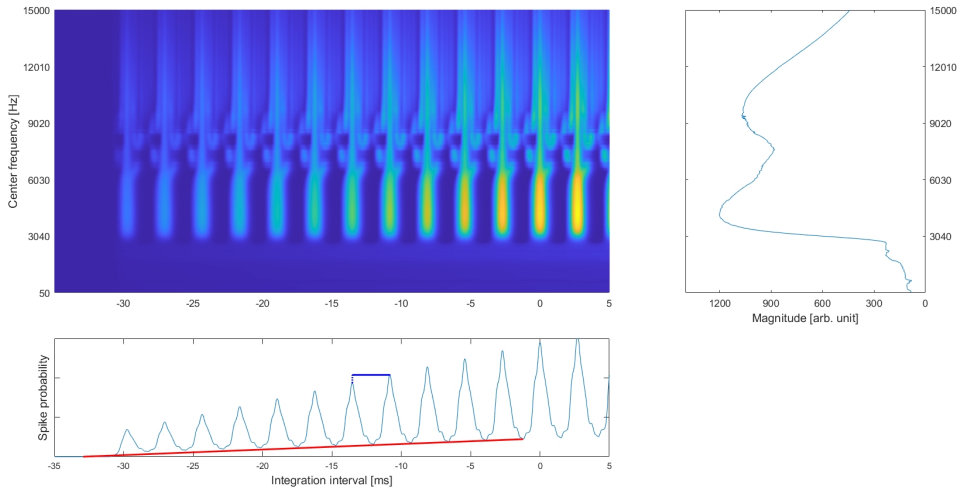


Figure 45: Stabilised auditory image of the clarinet tone F♯4. The upper left subfigure represents the neural activity in 500 basilar membrane regions or frequency channels (characterised by center frequency in [Hz]) per time in [ms]. Dark colours indicate low neural activity, bright colours indicate high neural activity. The upper right subfigure depicts the magnitude in arbitrary units per frequency channel. The lower subfigure shows the spiking probability across the temporal integration interval in [ms]. Note that the first 5 ms (i.e. from -35 to -30 ms) represent a baseline interval and the remaining 35 ms (i.e. from -30 to 5 ms) show the relevant data. The extracted sound features periodicity and hissiness are represented by the blue and red lines, respectively.

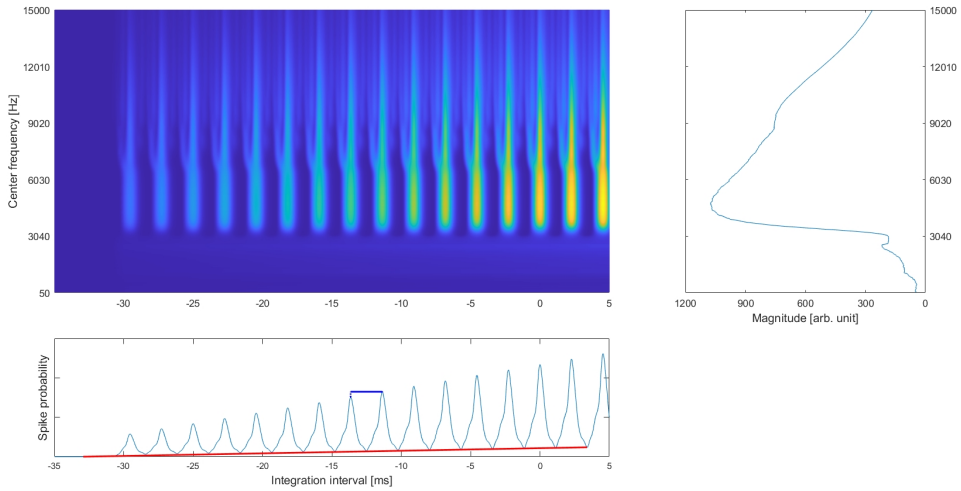


Figure 46: Stabilised auditory image of the clarinet tone A4. The upper left subfigure represents the neural activity in 500 basilar membrane regions or frequency channels (characterised by center frequency in [Hz]) per time in [ms]. Dark colours indicate low neural activity, bright colours indicate high neural activity. The upper right subfigure depicts the magnitude in arbitrary units per frequency channel. The lower subfigure shows the spiking probability across the temporal integration interval in [ms]. Note that the first 5 ms (i.e. from -35 to -30 ms) represent a baseline interval and the remaining 35 ms (i.e. from -30 to 5 ms) show the relevant data. The extracted sound features periodicity and hissiness are represented by the blue and red lines, respectively.

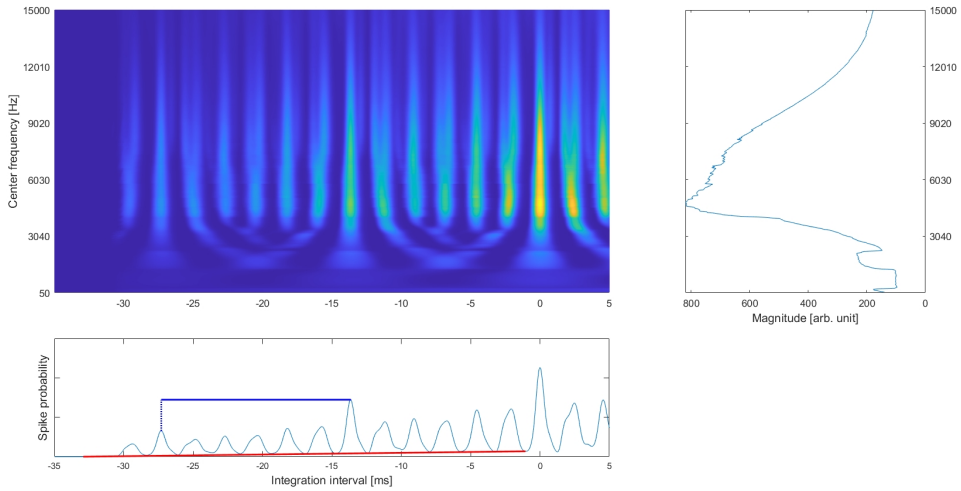


Figure 47: Stabilised auditory image of the bassoon tone D2. The upper left subfigure represents the neural activity in 500 basilar membrane regions or frequency channels (characterised by center frequency in [Hz]) per time in [ms]. Dark colours indicate low neural activity, bright colours indicate high neural activity. The upper right subfigure depicts the magnitude in arbitrary units per frequency channel. The lower subfigure shows the spiking probability across the temporal integration interval in [ms]. Note that the first 5 ms (i.e. from -35 to -30 ms) represent a baseline interval and the remaining 35 ms (i.e. from -30 to 5 ms) show the relevant data. The extracted sound features periodicity and hissiness are represented by the blue and red lines, respectively.

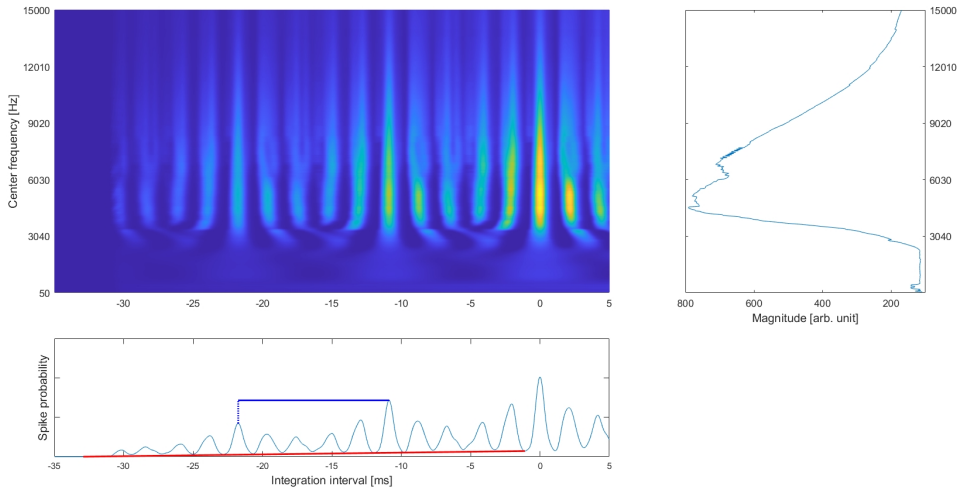


Figure 48: Stabilised auditory image of the bassoon tone F♯2. The upper left subfigure represents the neural activity in 500 basilar membrane regions or frequency channels (characterised by center frequency in [Hz]) per time in [ms]. Dark colours indicate low neural activity, bright colours indicate high neural activity. The upper right subfigure depicts the magnitude in arbitrary units per frequency channel. The lower subfigure shows the spiking probability across the temporal integration interval in [ms]. Note that the first 5 ms (i.e. from -35 to -30 ms) represent a baseline interval and the remaining 35 ms (i.e. from -30 to 5 ms) show the relevant data. The extracted sound features periodicity and hissiness are represented by the blue and red lines, respectively.

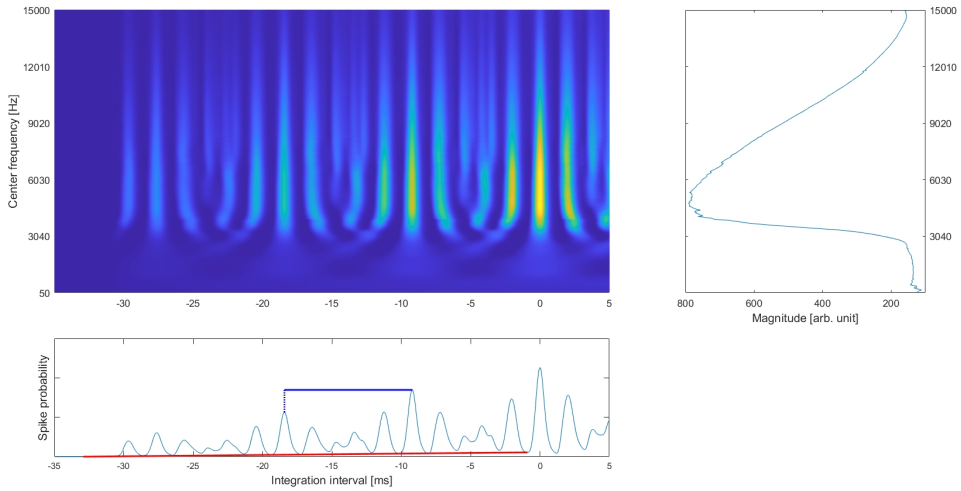


Figure 49: Stabilised auditory image of the bassoon tone A2. The upper left subfigure represents the neural activity in 500 basilar membrane regions or frequency channels (characterised by center frequency in [Hz]) per time in [ms]. Dark colours indicate low neural activity, bright colours indicate high neural activity. The upper right subfigure depicts the magnitude in arbitrary units per frequency channel. The lower subfigure shows the spiking probability across the temporal integration interval in [ms]. Note that the first 5 ms (i.e. from -35 to -30 ms) represent a baseline interval and the remaining 35 ms (i.e. from -30 to 5 ms) show the relevant data. The extracted sound features periodicity and hissiness are represented by the blue and red lines, respectively.

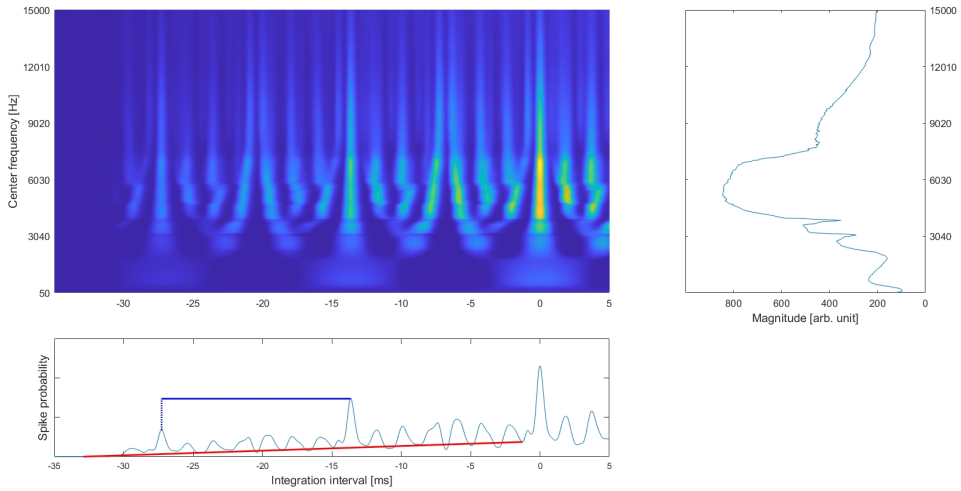


Figure 50: Stabilised auditory image of the bass clarinet tone D2. The upper left subfigure represents the neural activity in 500 basilar membrane regions or frequency channels (characterised by center frequency in [Hz]) per time in [ms]. Dark colours indicate low neural activity, bright colours indicate high neural activity. The upper right subfigure depicts the magnitude in arbitrary units per frequency channel. The lower subfigure shows the spiking probability across the temporal integration interval in [ms]. Note that the first 5 ms (i.e. from -35 to -30 ms) represent a baseline interval and the remaining 35 ms (i.e. from -30 to 5 ms) show the relevant data. The extracted sound features periodicity and hissiness are represented by the blue and red lines, respectively.

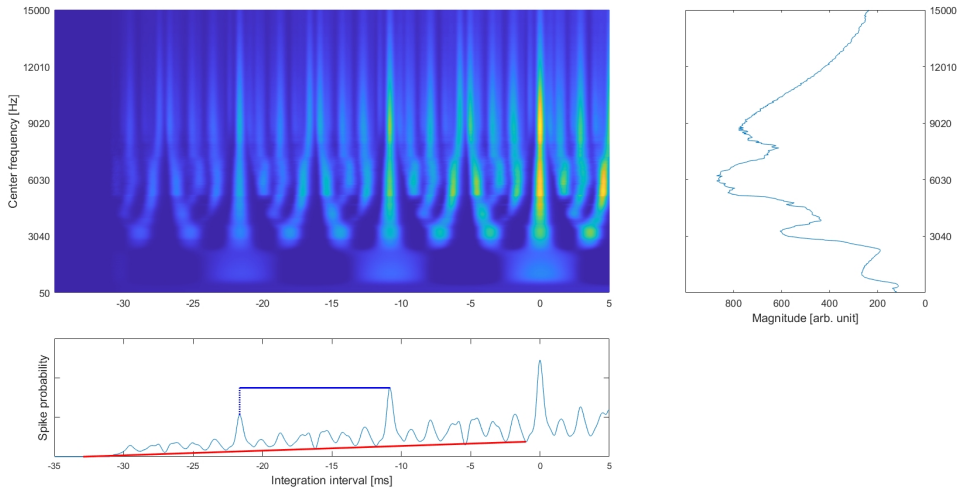


Figure 51: Stabilised auditory image of the bass clarinet tone F♯2. The upper left subfigure represents the neural activity in 500 basilar membrane regions or frequency channels (characterised by center frequency in [Hz]) per time in [ms]. Dark colours indicate low neural activity, bright colours indicate high neural activity. The upper right subfigure depicts the magnitude in arbitrary units per frequency channel. The lower subfigure shows the spiking probability across the temporal integration interval in [ms]. Note that the first 5 ms (i.e. from -35 to -30 ms) represent a baseline interval and the remaining 35 ms (i.e. from -30 to 5 ms) show the relevant data. The extracted sound features periodicity and hissiness are represented by the blue and red lines, respectively.

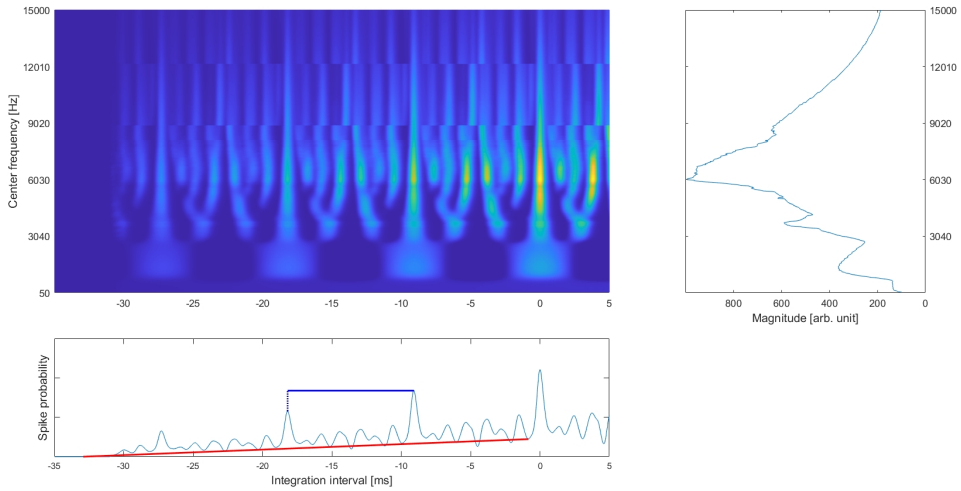


Figure 52: Stabilised auditory image of the bass clarinet tone A2. The upper left subfigure represents the neural activity in 500 basilar membrane regions or frequency channels (characterised by center frequency in [Hz]) per time in [ms]. Dark colours indicate low neural activity, bright colours indicate high neural activity. The upper right subfigure depicts the magnitude in arbitrary units per frequency channel. The lower subfigure shows the spiking probability across the temporal integration interval in [ms]. Note that the first 5 ms (i.e. from -35 to -30 ms) represent a baseline interval and the remaining 35 ms (i.e. from -30 to 5 ms) show the relevant data. The extracted sound features periodicity and hissiness are represented by the blue and red lines, respectively.

Acknowledgements

First and foremost, I would like to thank PD Dr. phil. André Rupp and Dr. Martin Andermann for assigning me the topic of this dissertation and the opportunity to work on it. Thank you for your friendly, reliable and motivating support, for your generosity and your patience. In this line, I would also like to thank Univ.-Prof. Dr. Christoph Reuter for the valuable contribution to this project.

I would like to extend particular gratitude to the Landesgraduiertenförderung of the University of Heidelberg for financing this PhD project via the individual doctoral fellowship of the LGF program.

I am grateful to Prof. Dr. med. Wolfgang Wick and Prof. Dr. med. Christoph Gumbinger for allowing me to work on the registry study FAST and for giving me the opportunity to gain experience in a medical work field. Special thanks go to Anja Ott, Dr. Loraine Busetto, Dr. Martin Andermann and Petra Günter for introducing me to the new task area and for the friendly and trusting cooperation.

I would like to thank especially Barbara Burghardt and Esther Tauberschmidt for their technical assistance and valuable support, and for their kind and appreciating manner. Furthermore, I want to thank my colleagues for the pleasant accompaniment and the many interesting, inspiring and encouraging conversations. Finally, I would like to thank all participants of this study for their time and effort.

Last but not least, I thank my partner, friends and family for their comprehensiveness, patience and support at all times.

Eidesstattliche Versicherung

1. Bei der eingereichten Dissertation zu dem Thema *Neuromagnetic representation of musical timbre dimensions in human auditory cortex* handelt es sich um meine eigenständig erbrachte Leistung.

2. Ich habe nur die angegebenen Quellen und Hilfsmittel benutzt und mich keiner unzulässigen Hilfe Dritter bedient. Insbesondere habe ich wörtlich oder sinngemäß aus anderen Werken übernommene Inhalte als solche kenntlich gemacht.

3. Die Arbeit oder Teile davon habe ich bislang nicht an einer Hochschule des In- oder Auslands als Bestandteil einer Prüfungs- oder Qualifikationsleistung vorgelegt.

4. Die Richtigkeit der vorstehenden Erklärungen bestätige ich.

5. Die Bedeutung der eidesstattlichen Versicherung und die strafrechtlichen Folgen einer unrichtigen oder unvollständigen eidesstattlichen Versicherung sind mir bekannt. Ich versichere an Eides statt, dass ich nach bestem Wissen die reine Wahrheit erklärt und nichts verschwiegen habe.

Heidelberg, 06.03.2024