Perception of Chord Components in Consonant and Dissonant Pure Tone and Piano Chords

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN
FACULTY OF SCIENCE
& MEDICINE

Thesis submitted in partial fulfillment
of the degree requirements

Master of Science

Author:
JOHANNA SALU
Graduate School of Neurociences
International Max-Planck Research School

Supervisor:
PROF. FELIX WICHMANN
Department of Computer Science
Eberhard-Karls Universität Tübingen

October 15th, 2016

Perception of Chord Components in Consonant and Dissonant Pure Tone and Piano Chords

---

Thesis
submitted in partial fulfilment of the degree requirements
Master of Science

Presented by:
Johanna Salu
from Tallinn, Estonia
Graduate Training Centre for Neuroscience (GTC),
Graduate School of Neural & Behavioural Sciences
Faculty of Science & Medicine
Eberhard-Karls Universität Tübingen

Thesis Advisor:
Prof. Felix Wichmann
Department of Computer Science,
Eberhard-Karls Universität Tübingen

Second reader:
Prof. Bettina Rolke
Department of Psychology,
Eberhard-Karls Universität Tübingen

Declaration of Authorship:

I hereby certify that this thesis report has been composed by me and is based on my own work, unless stated otherwise. No previous related work by myself has been reduplicated to contribute to this project. No other person's work has been used without due acknowledgement in this work. All references and verbatim extracts have been quoted, and all sources of information, including graphs, have been specifically acknowledged.

Johanna Salu   Tübingen, October 15, 2016

Acknowledgements

I would first like to thank my thesis supervisor, Prof. Felix Wichmann, for giving me a unique opportunity to work on this exciting research project - both a personal interest and a brainchild of mine. Despite it diverging from his usual research interests, his encouragement, enthusiasm and support from the very beginning guided this project to completion. The scientifically inquisitive and informed atmosphere in his lab made the experience of working on this project truly enjoyable.

I also want extend my thanks to the other members of the Neural Information Processing lab for the friendly atmosphere and stimulating lunchtime conversations. I want to specifically acknowledge Silke Gramer for her help with every administrative issue, Uli Wannek for his assistance in technical matters, Robert Geirhos for is helpful feedback on the first draft, David Janssen for his patience and help with programming questions, Tom Wallis (a former member of the lab) for his advice on experimental design and Heiko Schütt for his advice on data visualisation.

I would like to thank Barbara Peysakhovich for helping me find my way into Psychtoolbox, Caitlin Duncan and Judith Nieuwenhuis for their camaraderie and motivation during the final stretches of my Master thesis, and Yannik Bauer, without whom this thesis would not have looked half as nice, nor its author half as happy at the end of it. I'm very grateful for his help and keen eye in polishing the final product.

Finally, I would also like to thank Prof. Bettina Rolke for being the second reader of my thesis and the Graduate Training Centre of Neuroscience for providing a framework for this project.

Abstract

The use of simultaneously played tones, in the form of chords, is ubiquitous in Western music. Yet, we know little about how simultaneous tones are perceived – as a simple sum of their components or as a single sound with aggregate properties (McDermott and Oxenham, 2008). Previous work on chord perception with pure tone stimuli suggests that humans are unable to hear individual chord components when the number of components exceeds four (Demany and Ramos, 2005). It is well-known, on the other hand, that certain combinations of sounds result in pleasant and resolved, or consonant, chords and others in unpleasant and unresolved, or dissonant, chords.

In this thesis, I investigated how well individual chord components can be identified in five-tone chords consisting of pure and synthetic piano tones in a psychophysical yes-no detection task in which the consonance of the chords was varied. The consonant chords were built on the structure of a major triad and the dissonant chords were constructed of the most dissonant intervals used in Western music. Six subjects, comprising two professional musicians, three amateur musicians and one naïve listener, participated in the experiments.

In contrast to previous studies, we demonstrated that listeners were generally able to identify the components of consonant and dissonant chords, however, some components were easier to identify than others. Despite their higher complexity, components of piano chords were easier to identify, as well as components in certain absolute positions within the chord, notably the top and, to a lesser degree, the bottom note. Additionally, components with a strong harmonic resemblance and a functional relationship to the chord, such as root notes of the major triad, were detected better than other notes in the same absolute position. These components were also frequently misidentified as having been present. Musical training was found to improve subjects' performance, although large inter- and intra-individual differences were noted in different conditions. Our findings suggest that the harmonic relations between chord components play an important role in how they are perceived, and that sound complexity and musical training may contribute to highlighting these relationships.

Contents

Introduction

Music perception and language represent two uniquely human abilities.[1] Unlike language, which serves a clear practical purpose as a tool for communication and a vehicle for thought and cognitive function (e.g. Carmichael, Hogan, and Walter, 1932), no consensus exists on the evolutionary function of music.[2] Numerous evolutionary theories regarding the adaptive role of music have been put forward, ranging from music as a courtship display (Miller, 2000) to music as a facilitator of mother-infant interactions (Dissanayake, 2000). However, due to a paucity of empirical data, these *post hoc* theories remain difficult to falsify (McDermott, 2009) and consequently, lack predictive power. Nonetheless, despite having no immediately obvious adaptive function, creating and listening to music are present in virtually every culture around the world (McDermott, 2009).

While precise cultural practices vary, all music perception relies on certain perceptual and cognitive building blocks. Western musical tradition, in particular, builds upon abilities such as the perception of intervals (McDermott, Keebler, et al., 2010)[3] and melodic contours (Dowling and Fujitani, 1971; Edworthy, 1985) which give rise to relative pitch perception (Oxenham, 2012);[4,5] hierarchical beat perception which gives rise to musical meter (London, 2012);[6] and perception of consonance and dissonance which gives rise to tonality,[7] among others.

These perceptual abilities are not reserved for music but play a role in hearing in general. Moreover, these abilities tell us something about the underlying mechanisms of basic auditory processing. Music thus presents itself as a useful medium for probing general topics in auditory perception and cognition. It is not without problems, though, as the informational richness of musical stimuli makes them difficult to use in controlled experimental settings, yet it is equally difficult to distill out a single acoustic component while maintaining its musical relevance.

One central question in music perception relates to the processing of simultaneous tones. Unlike some other sound dimensions, such as loudness or brightness,[8] sound pitch can be used in a combinatorial way to generate novel perceptual experiences. One of the primary auditory functions, the grouping of simultaneous frequency components into one aggregate sound with a particular pitch and timbre[9] is a largely involuntary Gestalt-like perceptual process (Terhardt, 1974b). This allows us to perceive a single note played on a piano as one perceptual unit, and to differentiate it from a different note played on a piano, or from the same note sung by the human voice. On top of this basic level of frequency grouping, however, rests another layer of sound combination - the one at play when two notes on a piano are played simultaneously.

Western music, with the rare exception of Gregorian chants, takes full advantage of this combinatorial property of pitch to generate an array of new sounds. The use of multiple voices or different instruments is extremely commonplace. In their review, McDermott and Oxenham (2008), raise the question of what is actually perceived when several pitches are played at once - "multiple individual pitches" or a "single sound with aggregate properties". They note that while the

[1] Although see Hauser, Chomsky, and Fitch (2002), and Hauser and McDermott (2003).

[2] *"As neither the enjoyment nor the capacity of producing musical notes are faculties of the least use to man in reference to his daily habits of life, they must be ranked among the most mysterious with which he is endowed."* (Darwin, 1871, p.878).

[3] **Interval** - the distance in pitch between two notes (Sembos, 2006).

[4] **Pitch** - the perceptual correlate of periodicity (McDermott, 2013), and as a shorthand often described as the 'height' of a sound (Roederer, 2008).

[5] **Relative pitch** - the perception of pitches in relation to each other which allows listeners to recognise a melody irrespective of the absolute pitches played, so long as the relative relations between pitches are retained (McDermott and Oxenham, 2008).

[6] **Meter** - the orderly arrangement of weak and strong beats (Briggs, 2014), in layman's terms, what distinguishes a waltz from a march.

[7] **Tonality** - the set of rules used in the construction of a scale, as well as the central note in that scale (Sembos, 2006).

[8] **Brightness** - a key feature of timbre, determined by the spectral envelope of a sound (McDermott, Keebler, et al., 2010).

[9] **Timbre** - the quality of a sound which distinguishes between different instruments playing the same note (Roederer, 2008) and is determined by the relative amplitudes of harmonics in a complex tone McDermott and Oxenham (2008).

perception of polyphony[10] is interpreted in the framework of auditory scene analysis, the perception of homophony,[11] or chords[12], remains understudied, in particular with musical stimuli.

It was the aim of the present thesis to contribute to rectifying this apparent gap in our understanding of simultaneous tone perception. To this end, psychophysical experiments with musically relevant stimuli were conducted. A yes-no (single interval) detection task with chords composed of five tone components was used to probe how well listeners could identify the individual components of chords. The role of several factors, including sound complexity, chord consonance and the position of the component within the chord were explored. The following sections will cover the conceptual background and scientific literature pertaining to each of these topics, as well the motivation for the present study. I will begin, however, by providing an overview of the acoustic and music terminology and concepts that are relevant for understanding this work. Some of the terminology is explained in the sidenotes as it appears but a systematic summary of key concepts is provided in the next two sections.

[10] **Polyphony** - a musical texture in which the streams of sounds function relatively independently of each other (Briggs, 2014). The baroque fugues of Bach are a common example of polyphony.

[11] **Homophony** - musical texture in which the components move at a similar rhythmic pace (Language and Materials of Music), demonstrated by chord progressions.

[12] **Chord** - a set of three or more notes that are played simultaneously (Briggs, 2014); although some music theoreticians and scientists extend the term to include intervals.

Terminology and background

Acoustics

In this thesis, I will adopt the following definitions when describing the physical and perceptual properties of sounds.

Many sound pressure waveforms are produced by periodic physical processes, such as the vibration of human vocal chords, the oscillation of strings on a violin or the rotation of elements in a machine. The resulting waveforms are periodic - they consist of wave shapes that repeat at fixed intervals (McDermott, 2013).

Sound waves are usually characterised by their frequency - the rate at which the waveform repeats, expressed in Hertz (cycles per second). This frequency is known as the fundamental frequency, F_0 (McDermott, 2013). A tone consisting of a single sinusoidal waveform, known as a sine or pure tone, only has one frequency component. Tones with two or more constituent frequency components are known as complex tones (Hartmann, 2004).

Each of the individual sine waves comprising a complex tone is called a partial. Partials whose frequencies are positive integer multiples of F_0 are known as harmonic partials, or harmonics. Inharmonic partials are partials which deviate from a multiple relationship with the F_0. Most instruments produce some inharmonicity, notably percussion instruments such as timpani (some perceived pitch) or gongs (no perceived pitch). However, pitched instruments, such as violins or pianos[13] have partials that are close to the ideal harmonics (Hartmann, 2004). I will therefore use the term harmonic to refer to all frequency components of complex sounds in the remainder of the thesis.

[13] Pianos are actually tuned to be slightly inharmonic. The stretched tuning of pianos (i.e. an octave is greater than a factor of 2) is responsible for the 'warmth' of the piano timbre and electronic pianos need to duplicate this inharmonicity to reproduce its sound (Fletcher, Blackham, and Stratton, 1962). However, this type of inharmonicity does not affect the multiple relationships between partials (Hartmann, 2004).

A musical tone is defined as a sound which can be characterised by three features: pitch, loudness and timbre. It is our ability to assign these three qualities to a sound which distinguishes a tone from noise (Roederer, 2008). As noted above, the pitch of a tone is the perceptual correlate of the periodicity of the soundwave (McDermott, 2013), and usually a correlate of its fundamental frequency.[14] While humans can hear sounds with frequencies ranging from 20 to 20 000 Hz, only frequencies between 30 to 4000 Hz carry enough pitch information to convey melodic content (Oxenham, 2012).

[14] In truth, the definition of pitch is not as straightforward because pitch is also influenced by loudness and timbre (Roederer, 2008).

The pitch of a pure tone is equivalent to its frequency (Hartmann, 2004). A complex tone is perceived as having a pitch when its frequency components are harmonically related, or in other words, when they are harmonics. The perceived pitch then corresponds to the F_0 (McDermott, 2013). However, the fundamental frequency need not be present for listeners to perceive pitch, an effect known as the missing fundamental illusion (Licklider, 1954) in which sounds that lack the F_0 but whose remaining partials are multiples of a common frequency, still produce a perception of pitch.

Music theory

From a musical perspective, a tone may be described as a note. A note refers to a pitched sound in a musical context. In Western music, pitches with certain frequencies are assigned note names, for example the note with F_0 of 440 Hz, is known as A4. I will use the scientific pitch notation system to describe pitches which assigns a letter and a number to each note on a piano keyboard. Notes with F_0s in a ratio equal to an integer power of two are perceived to be very similar and are therefore assigned to the same pitch class (Roederer, 2008), resulting in the same letter name but different numbers. For example, a note with F_0 of 880 Hz, double of the F_0 of A4, is known as A5.

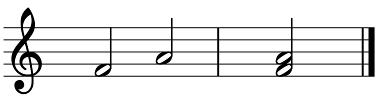

Figure 1: Major third as a melodic and harmonic interval.

In music, the relationship between the pitches of two notes is described as an interval. The interval between successive notes may be described as a horizontal, or melodic interval, whereas the interval between simultaneously played notes may be described as a vertical or harmonic interval (Briggs, 2014). See Figure 1 for illustration of interval notation. I will henceforth use the term 'interval' to denote exclusively the latter type of intervals. Moreover, I will occasionally use the term to denote the simultaneous presentation of two tones, as is commonly done in musical parlance.

Human perception of pitch is approximately logarithmic (Colwell, 2006), therefore intervals are defined by frequency ratios between notes, rather than by absolute frequency differences. For example, the interval produced by a doubling of frequency (ratio 2:1, e.g. from 100 to 200 Hz, or from 200 to 400 Hz) is known as an octave. The smallest interval used in Western music is a semitone, obtained by dividing the octave into 12 steps. A semitone is exemplified in the pitch difference between neighbouring keys on a piano. Each interval can then be described by the number of semitones it contains. The simple (i.e. sub-octave) intervals used in Western music, along with their frequency ratios according to just intonation[15] are shown in Figure 2.

[15] **Just intonation** - a system of tuning in which the pitches of a scale are determined by using intervals with small integer ratios.

Interval	Frequency ratio	Number of semitones
Minor second	16:15	1
Major second	9:8	2
Minor third	6:5	3
Major third	5:4	4
Perfect fourth	4:3	5
Tritone	45:32 or 64:45	6
Perfect fifth	3:2	7
Minor sixth	8:5	8
Major sixth	5:3	9
Minor seventh	16:9	10
Major seventh	15:8	11
Octave	2:1	12

Figure 2: Intervals and their associated frequency ratios and number of semitones according to just tuning. Two ratios are provided for the tritone because in just tuning the interval is different depending on whether it is treated as an augmented fourth or a diminished fifth.

Some intervals are defined by smaller integer ratios than others, such as the aptly named perfect fifth or perfect fourth. These intervals also tend to be perceived as the most pleasant and stable, or most consonant (Plomp and Levelt, 1965)(explained in more detail in section *Theories of Consonance*), and assigned higher ranks in tonal hierarchy (Malmberg, 1918). These intervals would also be used as the basis of just intonation. However, it is not possible to construct a 12-tone scale such that all possible intervals are consistently just (Hall, 1974), demonstrated, for example, by the syntonic comma.[16] drift sequence, illustrated in Figure 3. For example, if one were to play the sequence C G D A E C using the just intervals ratios for the ascending perfect fifth (3:2) and descending perfect fourth (3:4) and major third (4:5), then the last C of the sequence is higher than the first C by a ratio of 81:80, although it should be the same note.

[16] **Comma** - an undesirable microinterval resulting from different ways of tuning one note.

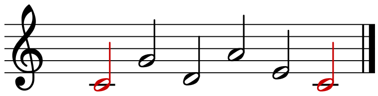

Figure 3: Demonstration of the inconsistency of just intonation. If the frequency ratios above are used, the first and last note in the sequence will be different by a frequency ratio of 81:80.

Different tuning systems, or temperaments, attempt to manipulate just intervals by small amounts in order to produce a consistent scale that does not result in any perceptibly out-of-tune intervals. The need for a good tuning is especially apparent with fixed-pitch instruments like the piano which cannot make on-line pitch adjustments in the same manner that a violin or the human voice can. Throughout history, different tuning systems have been used to divide the octave (Hartmann, 2004), the single interval consistent across the various tunings, into semitones, with varying degrees of success.[17] The standard temperament for fixed-pitch instruments in Western music today is equal temperament which divides the octave into 12 equal semitones (Hartmann, 2004). In equal temperament, the precise pitch does not affect the magnitude of the interval, allowing the use of key changes[18] in a piece at the cost of losing the small integer ratios for intervals. Equal temperament was used in this study because of its straightforward interpretation in different keys, as well as because one of the experiments involved piano tones.

[17] The quarter-comma meantone temperament, prominent in the 16th and 17th centuries resulted in one unusably wide fifth, called a 'wolf' (Page, 2004).

[18] **Key** - a set of notes that form a scale which determines the tonality of the piece.

Combining more than one interval results in a chord - a superposition of three or more notes. In this thesis, I will use the term 'component' in reference to the notes that comprise a chord or an interval. Different combinations of intervals give rise to distinct percepts and are used in different functions in music composition. These functions are driven by the tonality of the piece which sets up a system of hierarchical relations between pitches, and by extension, between chords composed of those pitches. However, a long musical progression is not necessary to establish a tonal context, and an isolated chord may already be enough for a listener to begin forming relations between the components.

Tonal context ascribes different roles for the components of a single chord. For example, in the major triad[19,20] composed of C4, E4 and G4, shown in Figure 4, the C is the root tone, or tonic - the most perceptually stable pitch, often perceived as the 'home' note and a pleasant final pitch in a musical line (Parncutt, 2011). When the C4 is transposed up in pitch by an octave, the root of the resulting chord E4, G4, C5 is still C. This second chord is an inversion of the first.[21] This simple example demonstrates that in music theory, and likely in perception, the relative position of a tone within a chord may play a more important role than its absolute position.

[19] **Major** - a characteristically happy-sounding musical scale, marked by the fact that the third note in the scale forms a major third interval with the root note.

[20] **Triad** - a chord consisting of three components which are arranged in thirds.

[21] **Inversion** - a chord shape in which the root is not the bottom note.

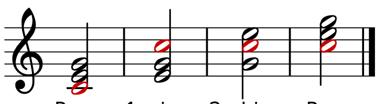

Figure 4: The major triad and its inversions. Inversions are created by transposing the lowest note up by an octave. The absolute position of the root note, C, changes across inversions. The third inversion is equivalent to the root inversion.

At the same time, using different absolute pitches for chords but retaining the pitch ratios gives rise to similar percepts, especially in equal temperament.[22] With this consideration in mind, the same chord structures in different keys were used in this study to create more perceptual variety while retaining a small number of structurally different stimuli.

[22] The famous collection by J.S. Bach, *The Well-Tempered Clavier*, contained pieces for 24 different keys. In his day, using the well-tempered tuning, each of the keys would have sounded slightly different, due to different interval sizes, resulting in a distinct 'colour' for each key (Page, 2004).

Complexity of sound

The simplest tone is a pure tone (described in *Terminology and background: Acoustics*). Since pure tones only consist of a single frequency, linear operations do not change the wave shape of the tone (Hartmann, 2004). Pure tones are virtually nonexistent in nature[23] but ubiquitous in hearing research. Due to the physical properties of natural objects, natural sounds consist of a superposition of sinusoids with different frequencies - complex tones.

[23] Although soft whistling and sound produced by tuning forks come close (Ballora, 2003).

In large part owing to their simplicity, pure tone stimuli are extensively used in acoustic research. A quick search in the database of the Journal of the Acoustic Society of America, a leading journal in the field, reveals that entries pertaining to pure tones outnumber those pertaining to complex tones by a ratio of 17:10. A likely assumption underlying the use of pure tones, but not specific to the field of psychoacoustics, is that performance on most tasks is highest with the simplest possible stimuli. Complex stimuli impose a higher informational load on the brain which requires increasingly more cognitive elaboration, sacrificing processing speed and accuracy. This principle is demonstrated in a wide range of perceptual (e.g. Bitterman, Krauskopf, and Hochberg, 1954; Heyer, Ryan, and MacDonald, 1976) and cognitive (e.g. Bethell-Fox and Shepard, 1988; Alvarez and Cavanagh, 2004) tasks.

According to this assumption, one might expect the detection of chord components to be best with pure tone chords because pure tones minimise the amount of sensory information that needs to be deliberated on any trial. If a chord were composed of three frequencies of equal amplitude, one would need only to process and store these frequencies to perform well in a detection task. On the other hand, if the chord were composed of three complex tones with a large number of constituent frequencies of different amplitudes, one would first need to parse the array of frequencies to identify the three components and commit significantly more information to memory. This assumption is supported by the observation that the pitch of the F_0 of a complex tone is sometimes confused with another pitch, usually an octave or a perfect fifth different from the true F_0 (Terhardt, 1974b).

In certain contexts, however, additional stimulus features might facilitate perceptual processing. First, if increased complexity were associated with familiarity, it is plausible that stimulus discrimination would improve through some fine-tuning of the perceptual system. It is well documented that in the visual domain, certain complex stimuli (e.g. faces) are processed in a more efficient manner than unfamiliar stimuli of equivalent visual complexity (Homa, Haver, and Schwartz, 1976). If this were the case, performance would improve if familiar piano sounds were used instead of the highly unnatural pure tones.

Secondly, certain tasks may benefit from additional information if that information provides a new strategy for performing the task. An obvious example for such a task is the pop-out strategy of visual search (Treisman, 1985). Perhaps the harmonic structures of complex sounds provide information that, rather than clutter and complicate the picture, would provide helpful cues for solving the task. It has been shown that frequency discrimination is superior with complex tones (Zeitlin, 1964), indicating that the presence of a harmonic structure can be beneficial in certain auditory tasks. This could be especially true for musically trained listeners who are used to drawing upon the full array of available musical cues in order to perfect their performance. This possibility is remarked by Roederer (p. 5): "A highly trained musician may have greatest difficulty in matching the exact pitch of a single electronically generated tone deprived of upper harmonics, fed to her ears through headphones, because her central nervous system is lacking some key additional information that normally comes with the real sounds with which she is familiar." Additionally, McDermott, Lehr, and Oxenham (2010) showed that the amount of musical training is positively correlated with listeners' preference for harmonically related frequencies in non-musical stimuli.

In order to distinguish between these alternative hypotheses, the experiment was carried out with pure tone chord and with piano stimuli. Listeners with varying degrees of musical training were recruited to explore whether training bears any consequence on the perception of pure and complex tones.

Consonance and dissonance

Theories of consonance

The observation that some tone combinations sound more pleasing to the ear than others already piqued the interest of ancient Greek scholars such as Pythagoras who noted that the pleasant, consonant, intervals were produced by pitches whose frequencies formed small integer ratios (Plomp and Levelt, 1965) (see Figure 2). Nonetheless, the concept of consonance has remained rather vague (Kameoka and Kuriyagawa, 1969; Plomp and Levelt, 1965; Cazden, 1980), not the least because the terms 'consonance' and 'dissonance' are used both in the context of isolated chords and intervals (i.e. to describe how their components interact) and musical passages (i.e. to describe how well a certain note or chord fits into the harmonic progression). The former is sometimes known as sensory or tonal consonance, and the latter as musical consonance (Cazden, 1980). This thesis explores the former type of consonance.

Sensory consonance is usually roughly defined as the perceived pleasantness of a combination of tones, despite the inadequacy of this definition in accounting for the vagueness of the term 'pleasant'(Parncutt and Hair, 2011),[24] despite recent evidence suggesting that mild dissonance is actually preferred over consonance in isolated chords (Lahdelma and Eerola, 2016). It should also be noted that consonance is not a binary feature: ranking of intervals for consonance forms a hierarchy (meta-analysis by Schwartz, Howe, and Purves (2003)).

Throughout the 20^{th} century, dissonance was thought to arise due to the phenomenon of beating,[25], giving rise to the sound quality of roughness which listeners find unpleasant (Terhardt, 1974b). An illustration of beating in a dissonant, compared to a consonant chord is shown in Figure 5. This theory of consonance and dissonance is credited to Helmholtz,[26] and popularised by Plomp and Levelt (1965) who proposed that consonance arises because consonant intervals have fewer harmonics which interact on the basilar membrane and pass

[24] Indeed, it is rather unhelpful for finding an objective definition of consonance if investigations solely rely on the criterion of 'pleasantness' to categorise their stimuli, only to find in the end that 'pleasantness' is the best descriptor of consonance.

[25] **Beating** - rapid fluctuations in waveform amplitude caused by the interference between sine waves of different frequencies (McDermott, Keebler, et al., 2010).

[26] Although much older records of similar statements exist (see Plomp and Levelt, 1965).

through the same auditory filters (see also Hutchinson and Knopoff, 1978). This, however, does not explain why consonance and dissonance are still perceived on dichotic presentation of sound (e.g. Bidelman and Krishnan, 2009; McDermott, Lehr, and Oxenham, 2010).

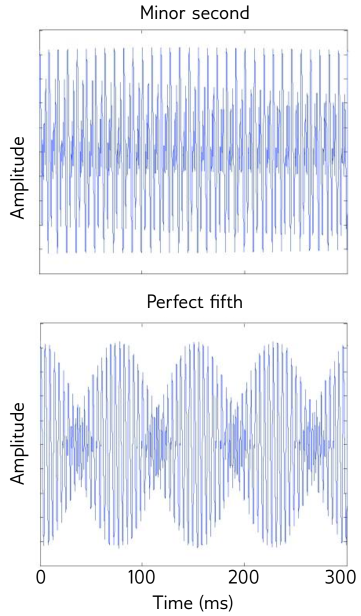

Figure 5: Beating is more pronounced in dissonant chords than in consonant chords. A minor second and a pure fifth are compared here. Figure adapted from McDermott, Lehr, and Oxenham (2010).

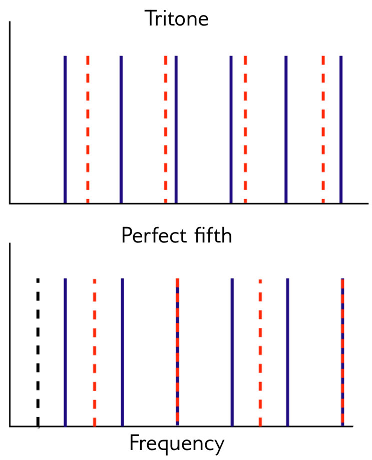

Figure 6: The harmonics of a perfect fifth consisting of an A (blue) and E (red) all align with the multiples of a common F_0 (black). The harmonics of a tritone consisting of an A (blue) and Eb (red) do not align. Figure adapted from Plack (2010).

Furthermore, as McDermott, Lehr, and Oxenham (2010) points out, consonant and dissonant intervals also differ in another respect, namely, how harmonically related the frequency components of the two constituent notes are. Ignoring the slight inharmonicities in real instrument sounds, the frequency components of two complex piano sounds forming a perfect fifth are all multiples of a common F_0, even if not every harmonic of that F_0 is present. A schematic representation of this relationship is shown in Figure 6. As such, consonant intervals closely resemble the harmonic spectra of single notes (McDermott, Lehr, and Oxenham, 2010). This resemblance is not present in dissonant intervals.

In a clever experiment capitalising on the individual differences between listeners consonance preferences, McDermott, Lehr, and Oxenham (2010) showed that listeners' consonance preferences were consistently correlated with their preferences for harmonicity in non-musical stimuli but not with their preferences for non-beating sounds. Additionally, Cousineau, McDermott, and Peretz (2012) found that while amusic listeners showed no preference for consonant chords, nor harmonicity, they exhibited normal preferences for non-beating sounds. These findings suggest that rather than deriving their pleasantness from a lack of beating, consonant intervals derive their pleasantness from their resemblance to the harmonic spectrum of a single note.

The third major theory to explain consonance describes it as a product of enculturation and familiarity, rather than an acoustic-physiological phenomenon (e.g. Lundin, 1947; McLachlan et al., 2013). This approach gains credence from the observation that equal temperament, omnipresent in Western music, results in less than perfectly harmonic intervals, and that certain tribes unexposed to Western music rate consonant and dissonant chords as equally pleasant, while fully capable of discriminating between the two (McDermott, Schultz, et al., 2016). However, this theory of consonance relates more to the aesthetic aspect of consonance, not to the basic perceptual processes that allow us to distinguish between these two types of sounds. The role of enculturation in consonance is further challenged by the findings that infants can distinguish consonant intervals from dissonant ones (Schellenberg and Trainor, 1996), and that the hierarchical ranking of intervals according to consonance can be predicted by a cat auditory nerve model (Bidelman and Heinz, 2011). In neither of these circumstances should enculturation play a large role (although the extent of prenatal and early life exposure to sounds is always a source of uncertainty in infant studies), suggesting that the perception of consonance is, at least partly, determined by physical properties of the sound. As our research was concerned with discrimination abilities, rather than with aesthetic responses as a function of consonance, the harmonicity hypothesis was adopted for this work.

Consonance and the perception of simultaneous tones

While most of the theoretical and practical work on consonance has been done using intervals, similar basic principles should translate to the consonance and dissonance of chords.

Consonant intervals are more frequently misidentified as a single tone than dissonant intervals (Dewitt and Crowder, 1987). This implies that consonant tones are more often fused into a single sound and should, consequently, be more difficult to identify as individual pitch components.

If consonant chords have a tendency to be perceptually fused, what percept do they fuse into? Researchers have attempted to determine the most salient frequencies in intervals by measuring the frequency following responses (FFR) of the brain stem, evoked potentials generated by periodic auditory stimuli. In one experiment, subjects listened to consonant and dissonant intervals, each built on the lower frequency of 220 Hz (Bidelman and Krishnan, 2009). The most salient pitch was 220 Hz which Bidelman and Krishnan labelled as the 'heard pitch'. Consonant intervals produced more robust neural responses to this pitch than dissonant intervals, and the magnitude of the response followed the consonance hierarchy of intervals. In accordance with the harmonicity hypothesis of consonance, the FFR responses to consonant intervals peaked at the harmonic multiples of the 220 Hz frequency, possibly reinforcing the prominence of this pitch. These findings thus suggest that the dominant percept in consonant chords is the bottom component of the interval. This is in line with the convention in music theory that the pitch determining the key of a chord is found in the bottom position in the root chord form. Likewise, the bassline is often used to play the root notes of chords in music.

Another FFR study where the top note remained the same across two intervals (minor and major seventh), showed that while the frequency of the bottom note dominates the brainstem response, corroborating Bidelman and Krishnan's findings, musicians in particular have a heightened response to the upper harmonics of the top note (Lee et al., 2009). The authors suggest that musical training improves listeners' ability to attend to the upper voice, given that melody is usually carried in the upper voice. Indeed, mismatch negativity studies suggest that changes in the upper voice are most easily detectable (Fujioka et al., 2005). Taken together, these two studies suggest that although musical experience can fine-tune auditory processing to attend more to musically relevant frequency components, the automatic auditory response emphasises the lower interval component.

Both studies appear to support the hypothesis that consonant tones are more susceptible to tonal fusion (Dewitt and Crowder, 1987) and that consonant intervals reinforce a certain perceived pitch which aligns with their combined harmonic spectrum. This suggests that the components of dissonant intervals should be more easily discriminated when played simultaneously. These studies, however, did not measure the actual discrimination of the chord components in a psychophysical task, and the relationship between subcortical activity and perception need not be straightforward. A further limitation of these studies is the use of the same bottom (Bidelman and Krishnan, 2009) or top note

(Lee et al., 2009) for all the intervals, possibly introducing a priming confound into their pitch salience analyses.

A discrimination task directly probing listeners' ability to identify the components of consonant and dissonant piano chords was performed by Seror and Neill (2015). In this experiment, listeners were instructed to respond whether an separately played probe note (e.g. C4) was contained in a subsequent interval or a semitone different from one of the interval components. The interval was either consonant (e.g. C4 and G4) or dissonant (e.g. C4 and F#4), and the probe targeted either the top or the bottom component of the interval. The authors reported that the component was identified faster and more accurately in consonant chords than dissonant chords but only in the bottom position of the interval. No consonance advantage was found for the top note. By demonstrating a consonance advantage, these findings go against the tonal fusion approach to interval (and chord) perception. More interestingly, however, they reveal an interaction between consonance and position in the perception of interval components which builds upon the harmonicity theory of consonance.

Consonance and component position

If consonance is governed by the similarity of the consonant sound to the harmonic profile of a single complex tone, it seems plausible that the percept of consonant sounds would, to a certain degree, resemble the percept of that single tone. Evidence from brainstem recordings suggests that consonant intervals do elicit responses that peak at the harmonic multiples of a F_0, namely the bottom tone in an interval. This would imply that, although consonant intervals are perceptually fused, this fusion reinforces one frequency component and actually facilitates the perception of the lower component in an interval. This would explain the findings of Seror and Neill (2015) who demonstrated that consonance can facilitate the identification of interval components but only in the bottom interval position.

Dissonant intervals, on the other hand, do not form harmonic series that resemble or reinforce that of any single pitch. This makes them unlikely candidates for tonal fusion, which would suggest an advantage for identifying the components of dissonant chords. Such an advantage, however, was not found (Seror and Neill, 2015). It is unclear whether the ability to simply detect the presence of more than one pitch, facilitated by the irregular harmonic profile of dissonant intervals, should translate into an advantage in identifying these components. Seror and Neill suggest that the correct identification of the pitches in dissonant intervals may actually be impaired by the irregularity of the harmonic spectrum which fails to clearly represent either of the components.

The interaction between consonance and component position is difficult to fully explore in the limited context of intervals which contain only two positions. Chords, on the other hand, present a set of component positions, equal to the number of components, and through the use of inversions of the same chord shape, the absolute position of a component could be varied independently from its functional position (determined by music theory and the harmonic spectrum of the chord).

These advantages motivated the use of chord stimuli in this thesis.

As previously mentioned, the topic of chord perception remains understudied. Several studies have assessed aesthetic responses (e.g. Lahdelma and Eerola, 2014) and consonance judgements in response to different chords in isolation (e.g. Rasmussen, Santurette, and MacDonald, 2014) and in harmonic progression (e.g. Bigand, Parncutt, and Lerdahl, 1996), and similarity judgements between different chord structures (e.g. Beal, 1981; Crowder, 1985; Gibson, 1986). Two experiments specifically investigating the perception of chords components (Demany and Ramos, 2005; Schnuerch et al., 2014) suggest that humans are incapable of detecting individual chord components when the number of components is more than four.

Musical training often requires students to learn to recognise common chord shapes (such as the major triad or dominant seventh), usually consisting of 3-4 components, by ear. Therefore, the ability to recognise the components of these chords, especially in a musically trained population, would not be surprising. However, given the nature of the training, it would then be difficult to determine whether a bottom-up perceptual or top-down cognitive strategy (e.g. retrieving the components that are consciously associated with the composite sound of common chord shapes) is at play. Therefore, the use of at least five chord components is recommended for psychoacoustic research.

Demany and Ramos (2005) investigated the ability to hear out the components of inharmonic (randomly composed) pure tone chords consisting of five components, and found that listeners could not reliably identify the chord components. More recently, Schnuerch et al. (2014) demonstrated change-deafness with musical (harmonic) pure tone stimuli comprising five and six components, so long as the time interval between successive chords was longer than60 ms, further supporting the view that chord stimuli, musical or not, cannot be reliably discriminated on the basis of their components.

Since these two studies used either non-musical stimulus structures, and/or non-musical tones, we set out to determine if there were any stimulus configurations under which chord components could be identified above chance.

Experimental questions and design

The aim of this thesis was to contribute to our understanding of how chords are perceived by investigating how well individual components of chords can be discriminated in a yes-no detection task. A yes-no task was chosen over a forced choice design because the former is more direct in probing the presence of individual components, rather than an overall change in the global chord percept. In a two-alternative forced choice (2AFC) task, both temporal intervals would contain a chord, one with the signal (component of interest) present and one with the signal absent. However, given how little we know about the perception of chord components, the choice of what chords to present in the signal absent intervals could have affected performance in unforeseen ways. Furthermore, a 2AFC design would not have allowed us to explore the false alarm rates with consonant incorrect components as in the yes-no design.

In order to determine how the consonance of the chord and the position of a component within a chord affect its detectability, we used five-component consonant and dissonant chords. Although consonance is not binary, we chose to represent the extreme ends of the continuum by using the major triad as the basis of the consonant chords, and the most dissonant intervals (tritone, seconds and sevenths) as the basis of the dissonant chords. Five inversions of a root chord shape (root shape included) were used in either category. The incorrect probes presented alongside consonant chords comprised both tones consonant with the chord, and dissonant tones. The incorrect probes presented with dissonant chords were all dissonant.

The relatively large number of components was to ensure that the stimuli would be sufficiently unfamiliar that listeners had to rely on perceptual, rather than high-level cognitive strategies or prior knowledge. Additionally, the choice to use five-component chords allowed us to construct five inversions of each chord in order to disentangle the factors of absolute and functional position in consonant chords. In consonant chords, constructed from the pitches of the major triad, the root note has a special functional role. Moreover, the harmonic spectra of the consonant chords most closely resemble that of the root note. As the absolute position of the root note varies across the inversions, we can distinguish between the absolute and functional position of a component.

According to the harmonicity hypothesis of consonance, and the literature on consonant and dissonant intervals, we expected the perception of consonant chord components to be affected by its position. Components corresponding to the root tone of the chord should be reinforced by the composite harmonic spectrum and therefore easier to detect. Furthermore, we expected that certain incorrect components which align well with the composite harmonic spectrum of the chord would be mistakenly identified as present. We did not expect there to be any systematic patterns in how well the different components of dissonant would be detected. Furthermore, it was unclear whether performance with dissonant chord components should be on par with that of the non-reinforced consonant components or above it. Additionally, there may be some advantage provided by certain absolute positions, notably the top and bottom positions which have a special status in music.

We also studied the effect of sound complexity on chord perception by using pure and piano tones. We predicted higher performance with piano sounds as we expected the harmonicity of the chord to facilitate perception, and piano sounds, unlike pure tones, provide harmonic information to the listener.

Finally, we tested subjects with varying degrees of musical experience to explore how these factors interact with musical training. Although we predicted professional musicians to generally outperform the subjects with less training, we also expected them to be more affected by the variations in sound complexity, position and consonance.

Methods

Participants

6 listeners (3 female) with a mean age of 27.8 ± 9.8 years (mean$\pm$ s.d.) participated the experiments. 2 listeners were professional musicians (mean age 39 ± 9.9 years), 3 listeners (mean age 22.3 ± 2.3 years) were recruited from with the prerequisite of having some musical training, grouped as amateur musicians, 1 naïve listener (age 22 years) was recruited without musical training prerequisites. The subjects' musical training was evaluated with a self-report questionnaire. The average number of years spent regularly performing music among the professional musicians was 30.5 ± 13.4 years. Among amateur musicians, the same measure was 11 ± 4 years. The naïve subject had 3 years of musical experience. The questionnaire may be found in Appendix A. None of the subjects had perfect pitch, any form of sound synaesthesia, nor any known hearing impairment.

Different probe amplitudes were used in the *piano tone* condition for the professional, and amateur and naïve subjects, explained below and in Appendix C. One additional subject (S1), who was the author (age 24 years, 17 of musical experience) completed the experiment, performing the *piano tone* condition twice, once with either amplitude setting to assess the effect of probe amplitude on task performance. Furthermore, subject S1 completed 2 additional experimental sessions with different timing parameters during the piloting of the study. All subjects except for S1 were naïve with respect to the aim of the experiment. The data from S1 is not reported in the Results but it was used to validate the comparison between the two different amplitude settings, shown in Appendix C.

Apparatus

Stimuli were presented binaurally through Ultrasone Signature PRO over-ear headphones with S-Logic PLUS technology (Ultrasone, Germany) while participants were seated in a quiet and dim room. The closed-back headphones provide up to 40 dB external noise attenuation, depending on sound frequency (*Headphone Measurements* 2011).

The analog signal to the headphones was generated with an 24-bit external sound card and headphone amplifier, Grace m903 (Grace Design, USA). Digital signal generation, output and response collection were implemented in MATLAB (2013b and 2016b) with the Psychtoolbox-3 extension (Kleiner et al., 2007), running on a Linux-powered desktop computer (Ubuntu 14.04) and laptop (Linux Mint 17.1).

Stimuli

Conforming to the yes-no task design, the stimuli comprised chords and probe tones. A Gaussian white noise mask was used between trials to minimise the influence of successive trials on each other. The stimuli in the *piano tone* and *pure tone* conditions varied in sound complexity. The compositional structures of the chords and probes used in both conditions were identical and will be described first.

Chords. The chords used in both experiments consisted of five frequency components. Every chord spanned a range of between 2-3 octaves and the interval between neighbouring tones was always below an octave.

Half of the presented chords were consonant, composed of the notes of a major triad. Figure 7 shows the consonant chord shapes used, with C3 as an example starting pitch, as well the probes associated with each chord shape. Additionally, four other starting pitches were used (G#2, A#2, D3 and E3). Chords Cons. 1 - Cons. 5 were functional inversions of each other. To contain the full range of frequencies required by all inversions to the middle five octaves of a piano, inversions were transposed to begin less than a perfect fifth above or below the starting pitch of the root inversion, in this case, C3. This transposition does not affect the relative relations, nor the absolute positions, of chord components.

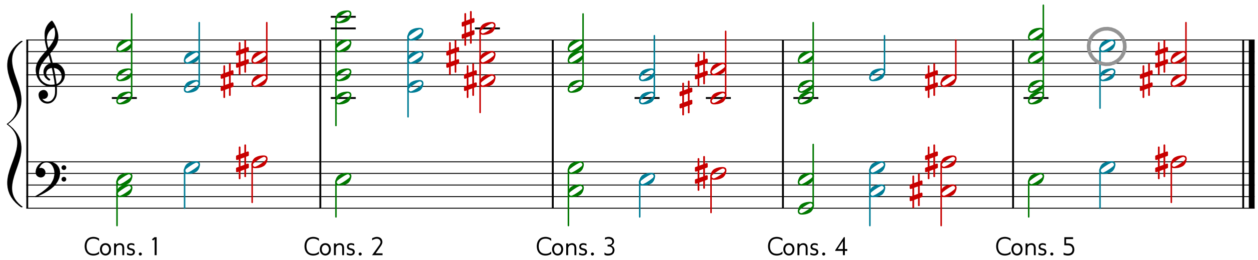

Figure 7: The consonant chord shapes and probes used, using C3 as the starting note. Green notes represent the chord (when played simultaneously) and correct probes (when played individually), blue notes represent consonant incorrect probes and red notes represent dissonant incorrect probes. Due to an error in the stimulus code the highest incorrect probe for Cons. 5 (indicated by grey circle) was replaced with a correct probe (C5). This component is appropriately addressed in the results.

The remaining half of chords were dissonant, composed of the tritone, major seventh, and major second intervals. Figure 8 shows the dissonant chord shapes used, again with C3 as an example starting pitch.[27] The same five starting pitches were used as with consonant chords (G#2, A#2, C3, D3, E3). Chords Diss. 1 - Diss. 5 are functional inversions of each other. Similarly to consonant chords, the inversions were transposed to begin less than a perfect fifth above or below the starting pitch of the root inversion.

[27] Six different dissonant probes were planned for each dissonant chord to match number of probes associated with each consonant chord, however, due to a coding error, only the three probes shown in Figure 8 were used in the final experiment. Furthermore, the incorrect probes for Diss. 5 with the highest starting pitch (E3) were inadvertently transposed down by a semi-tone, remaining equally dissonant but different in relative pitch from the other starting notes.

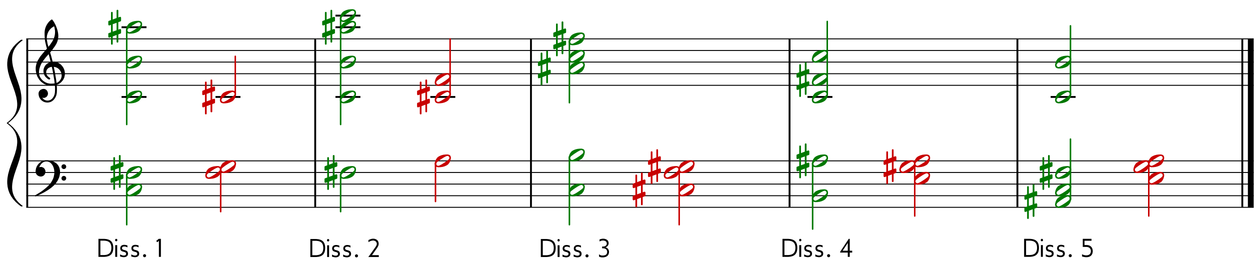

Figure 8: The dissonant chord shapes and probes used, using C3 as the starting note. Green notes represent the chord (when played simultaneously) and correct probes (when played individually), red notes represent dissonant incorrect probes.

As 10 different chord shapes were used (5 consonant, 5 dissonant), with 5 starting pitches each, a total of 50 different chord stimuli were used. However, for the purpose of our analyses, the same chord shapes with different starting pitches were treated as *functional repetitions.*

Probe tones. The probes in both experiments consisted of single tones. 11 different probe tones were associated with each consonant chord shape, 5 of which were correct probes (i.e. identical to the components that were used to construct the chord) and 6 of which were incorrect probes (i.e. tones which were not present in the chord). Similarly, 5 correct probes and 3 incorrect probes (each presented twice as often as each of the incorrect probes for consonant chords) were associated with each dissonant chord shape. The classification of probes is illustrated in the bottom panel of Figure 9. All probes were contained within the range of the chord such that no probes were higher or lower than the top or bottom tone in the corresponding chord, respectively.

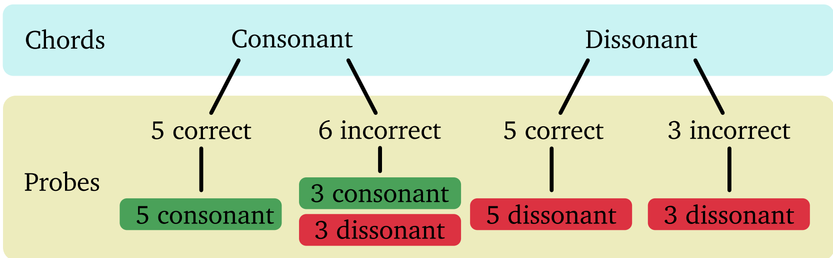

Figure 9: Organisation of probe tones. Consonant and dissonant chords were associated with 5 correct probes (the components of the chord itself), and 6 (consonant chords) or 3 (dissonant chords) incorrect probes. Half of the incorrect probes for consonant chords were consonant, all the incorrect probes for dissonant chords were dissonant.

Probes for consonant chords. Since the correct probes for consonant chords were identical to the chord components, these probes were considered consonant with respect to the chord. The incorrect probes for consonant chords were divided into 3 consonant and 3 dissonant probes. The incorrect consonant probes consisted of tones which were part of the same major triad that was used to construct the chord but which were not present in the chord. Each consonant chord was constructed to contain 3 consonant 'slots' of unused consonant tones, and these tones formed the set of incorrect consonant probes. The intervals between the consonant probes and their neighbouring chord components comprised consonant intervals - perfect fourths, a major thirds and a minor thirds. The incorrect dissonant probes were offset from the frequencies of the consonant probes so as to maximise dissonance with the chord. The dissonant probe formed either a tritone, minor second or a major second with at least one of its two nearest chord components.

Probes for dissonant chords. The correct probes for dissonant chords were identical to the chord components, however, due to the dissonant intervals formed between each of the neighbouring chord components, these probes were still considered to be dissonant. All incorrect probes for dissonant chords were similarly considered to be dissonant as the dissonant context established by the chord itself could not be resolved with any single tone. The incorrect probes always formed either a tritone, minor second, major second or a minor third with at least one of its two nearest chord components. For more information on the selection of chords and probes, see Appendix B.

Stimuli in pure tone condition. The stimuli used in the *pure tone* condition were generated with MATLAB. The chords consisted of 5 superimposed sine wave components of equal amplitude. The probes consisted of single sine wave tones, equal in amplitude to the individual sine wave components that comprised the chords. All chords and probes were windowed with a symmetric tapered cosine window (ratio of cosine-tapered section length to the lengtt of the entire window, $\alpha = 0.75$; an α of 0 would result in a rectangular window, an α of 1 would result in a Hann window) to eliminate clicking noise at the onset and offset of sound.

Stimuli in piano tone condition The chords and probe tones in the *piano tone* condition consisted of artificially generated piano tones. The stimuli were created with the music composition software MuseScore (MuseScore 2.0.3, 2016). First the chords and probes were manually entered using the music notation features of the programme. Its inbuilt synthesizer was used to generate realistic instrumental playback. The stimuli were then exported from MuseScore in .wav format and further cropped and windowed using an asymmetric tapered cosine window in MATLAB in order to retain the characteristic piano attack (first half of the window used $\alpha = 0.1$, second half used $\alpha = 0.75$). The probes, contained in individual .wav files had greater amplitudes than the individual components in the chord .wav files in which five notes were created and exported together. This resulted in amplitude differences between the presentations of the same note in the a chord and in the probe. The naïve and amateur subjects were tested with these amplitude differences present. In the later testing sessions with professional musicians, the amplitude of the probes in the .wav file was reduced by dividing it by the number of chord components.[28] For more information on stimulus generation and presentation, as well as the performance of S1 with the two amplitude-configurations, see Appendix C.

[28] The additional subject, S1, completed the *piano tone* condition with both amplitude configurations and showed not substantial differences in performance with the two amplitude settings.

Design

The experiment followed a self-timed single interval detection (yes-no) paradigm where the signal was present (*yes* trials) or absent (*no* trials) with 50% probability. All subjects completed both conditions and the order of conditions was counterbalanced (3 subjects completed the *pure tone* condition first).

Each condition consisted of 3000 trials, divided across two sessions. The 3000 trials were divided into blocks of 200, such that the first session comprised 8 blocks (1600 trials) and the second session comprised 7 blocks (1400 trials).[29] Stimulus presentation was pseudo-random. The subjects could pause the stimulus presentation after every 40 trials, and were encouraged to take a break after every block. Feedback about response accuracy was provided after every 40 trials, and at the end of every block. Each session began with 40 practice trials. The total duration of the experiment depended on the response times of the subjects, as well as the duration of their breaks, and ranged from 5 to 8 hours.

[29] S7 completed the experiment in three sittings: session 1 and 3 blocks of session 2 of the *piano tone* condition in the first sitting, the remaining 4 blocks of session 2 of the *piano tone* condition and session 1 of the *pure tone* condition in the second sitting, and session 2 of the *pure tone* condition in the third sitting.

On a given trial, each of the 50 chord stimuli could be paired with one of the 11 probes (5 correct, 6 incorrect). However, in order to balance the number of 'yes' and 'no' trials, each chord stimulus was actually paired with 12 probes such that one of the correct probes was played twice. Each correct probe appeared equally many times as the 'extra' probe.

Task

Both experiments followed the same task structure. In every trial, subjects heard a chord, followed by a probe tone and were instructed to respond whether they thought that the probe had been present in the chord and indicate their decision with a key press.

The outline of a trial is shown in Figure 10. Each trial began with a 250 ms presentation of white noise followed by a 150 ms interval of silence. Next a chord was presented for 500 ms followed by another 150 ms interval of silence. Then a probe tone was presented for 500 ms, followed by a silent interval during which subjects were expected to make a response. After a response was made, the trial ended with a 250 ms presentation of visual feedback in the form of either a green (correct response) or a red screen (incorrect response). Simultaneously with visual feedback, white noise, signalling the start of the next trial, was presented.

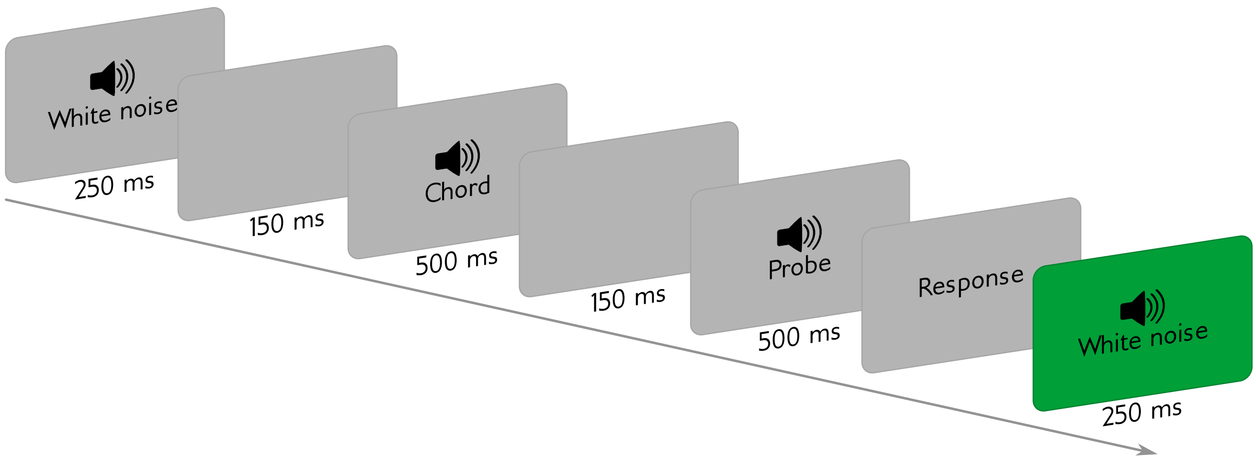

Figure 10: Outline of a single trial. The beginning of a trial is indicated with 250 ms burst of white noise which overlaps temporally with the visual feedback (red or green screen) from the previous trial. If there is no previous trial, as is the case for the first trial in a block, then the white noise is presented alongside a grey screen, as illustrated here.

Subjects were specifically instructed to avoid using any cognitive or musical strategies, such as whistling, humming or singing during the experiment and to respond as quickly as possible based on their initial impression, rather than attempt to recreate the chord in their mind. The first block (200 trials) of S2 was repeated because of the subject's use of an undesired strategy.

Results

The results report the response accuracies of subjects in the chord component detection task. Performance was explored on an individual subject level and due to time constraints, no statistical group analysis was conducted. All error bars show the 95% confidence interval of the normal approximation to the binomial distribution. Response times were collected but not reported here.

Starting notes

Firstly, to validate the comparison of chord structures collapsed across different starting notes, i.e. *functional repetitions*, the overall performance of each participant across the transpositions was plotted. Figure 11 shows the percent correct of each participant for the consonant and dissonant chord shapes for each of the five possible starting notes in the *pure tone* condition. Equivalent performance in the *piano tone* condition is shown in Figure 12.

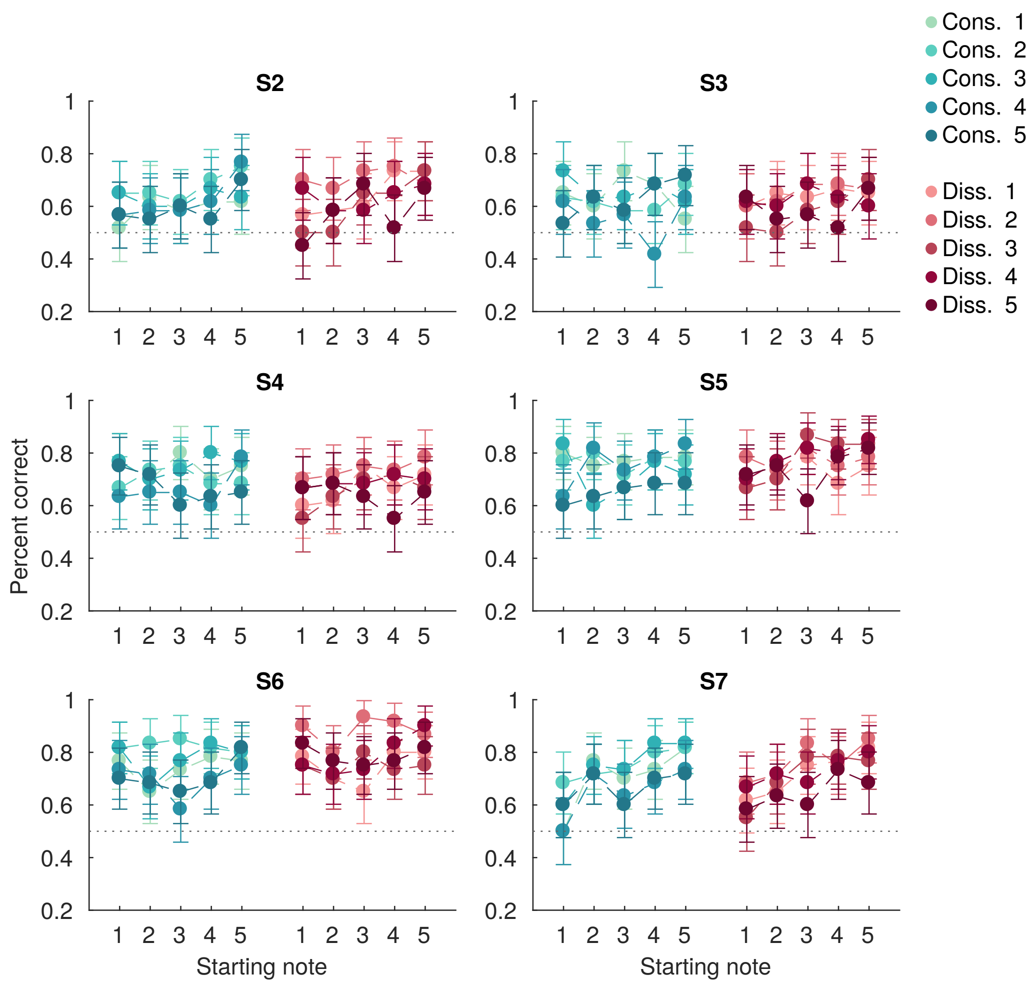

Figure 11: Performance across the five different starting notes for each chord shape in the *pure tone* condition. Performance of each subject is shown separately. Blue points represent different consonant chord shapes, red points represent dissonant chord shapes. Error bars show the 95% confidence interval of the normal approximation to the binomial distribution. Grey dotted line represents chance-level performance.

As seen in Figures 11 and 12, the performance of listeners across different starting notes varies slightly but largely unsystematically. Overall, the data validate our use of different starting notes for the same chord shapes as *functional repetitions*. In the *pure tone* condition, there is a slight upwards trend across the different starting notes, illustrated in particular by S7, demonstrating that the task was slightly easier with higher absolute pitches. This trend is not visible in the *piano tone* condition, with the exception of S2. In the *piano tone* condition, performance is more constant across the different starting pitches, with the exception of a large drop in performance with the highest starting note in the Diss. 5 chord for subjects S4, S5, S6 and S7. This drop is completely absent in the *pure tone* condition, in fact, there is an upwards trend in the corresponding chord Diss. 5 across the starting notes in the *pure tone* condition. It is likely caused by the inadvertent use of different dissonant probes in the Diss. 5 with starting note 5. There appears to be no systematic difference in the effect of starting note variation on consonant and dissonant chords.

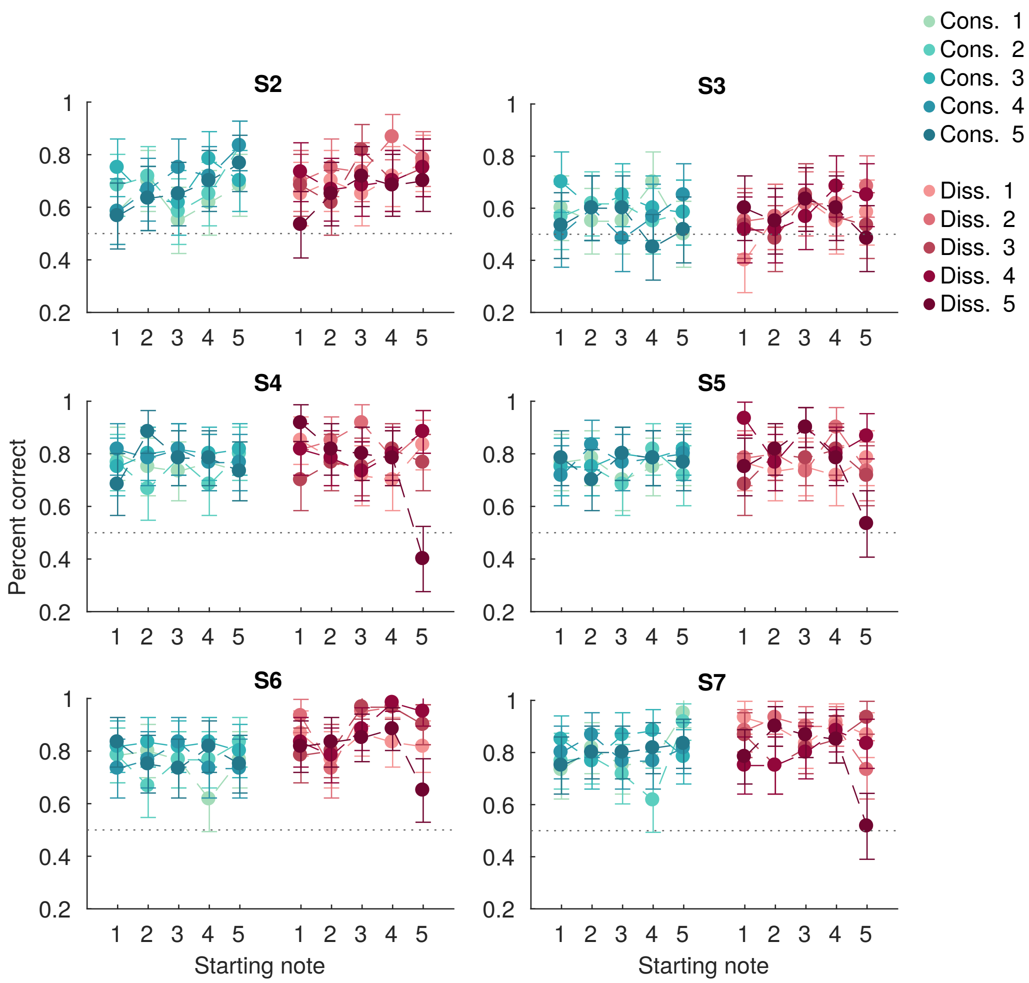

Figure 12: Performance across the five different starting notes for each chord shape in the *piano tone* condition. Performance of each subject is shown separately. Blue points represent different consonant chord shapes, red points represent dissonant chord shapes. Error bars show the 95% confidence interval of the normal approximation to the binomial distribution. Grey dotted line represents chance-level performance.

Sound complexity

The performance of subjects in the *pure tone* and *piano tone* conditions is shown in Figure 13. With the exception of S3, the naïve subject, all subjects performed better in the *piano tone* condition. The difference in performance between the two conditions varied noticeably between subjects, with the mean difference of all subjects at 4.2% higher in the *piano tone* condition. The greatest difference, 10.9%, was shown by S7.

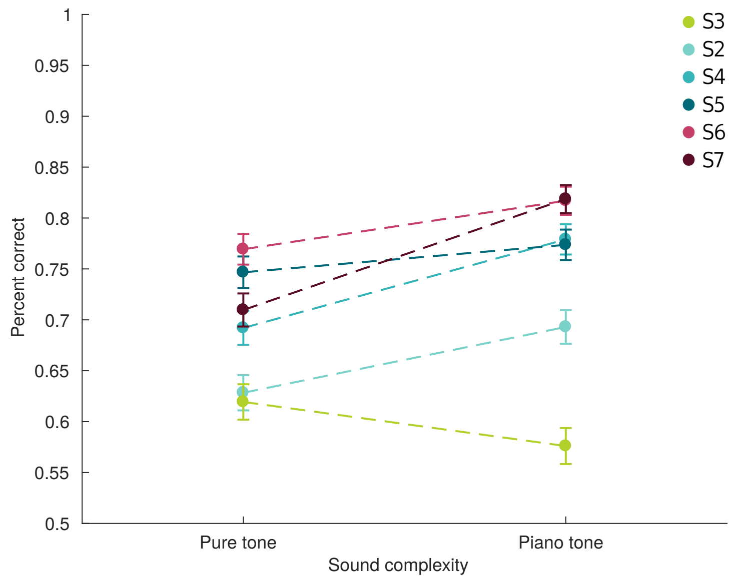

Figure 13: Performance in the *pure tone* and *piano tone* conditions, all chords combined. Error bars show the 95% confidence interval of the normal approximation to the binomial distribution.

In order to determine how the order of completing the two conditions and potential learning affected subjects' performance, the performance across the four experimental sessions, in chronological order, was examined. Figure 14 shows subjects' average performance in each of the four sessions (in colour), as well as a more detailed display of performance on every block of 200 trials (in grey). Subjects S2, S4 and S6 completed the *pure tone* condition first and their performance gradually increased in accuracy. A large increase in accuracy from session 3 to session 4 (*piano tone* condition) occurred for S2. S3, S5, and S7 completed the *piano tone* condition first. For S3, the naïve subject, performance accuracy increased from session 1 to session 2 (*piano tone* condition) but did not continue to improve, plateauing around 60%. For S5 and S7, performance also increased from session 1 to 2 (*piano tone* condition) but decreased in the *pure tone* conditions (sessions 3 and 4), showing a trend opposite to S2, S4 and S6.

Figure 14 clarifies why the performance S3 diverged from others in Figure 13. Unlike the more experienced listeners, S5 and S7, who underwent the same order of experimental sessions, S3 performed close to chance-level in their very first session. Figure 14 also demonstrates that the large difference between performance in the two conditions for S7 is largely due to their lower performance in the *pure tone* condition. For example, they are outperformed in the *pure tone* condition by the

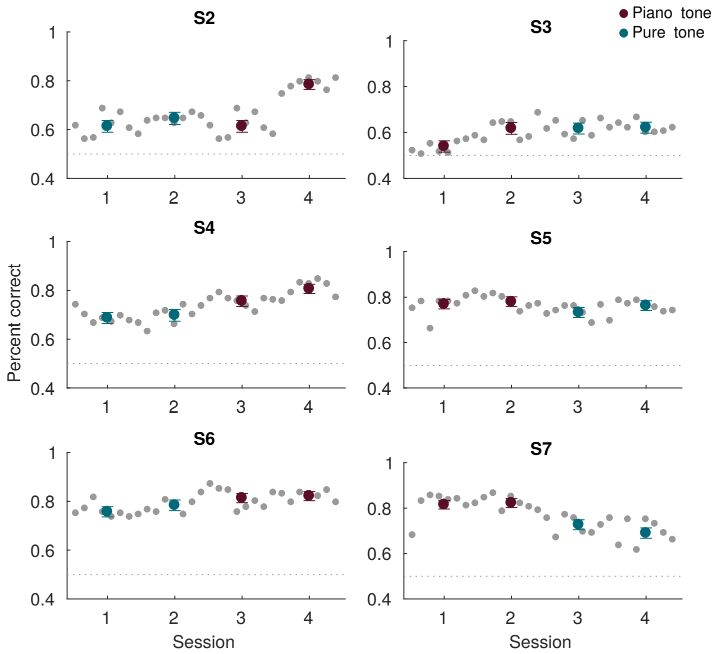

Figure 14: Performance across the four experimental sessions, ordered chronologically. S2, S4 and S6 completed the *pure tone* condition (indicated by blue points) first; S3, S5 and S7 completed the *piano tone* condition (indicated by red points) first. Grey points show performance on every block of 200 trials, ordered chronologically. Error bars show the 95% confidence interval of the normal approximation to the binomial distribution. Grey dotted line represents chance-level performance.

other professional subject S6 who completed the condition first, with less experience with the task and the chord structures.

Figures 13 and 14 also demonstrate that performance correlated roughly with musical training. The naïve subject, S2 (3 years of musical experience), had the lowest overall performance (59.8%), followed by S3 (7 years, 66.1%), S4 (11 years, 73.6%), S5 (15 years, 76.0%), S7 (21 years, 76.4%) and S6 (40 years, 79.3%). The large difference in the performance of S7 in the two conditions suggests that musical experience may interact with sound complexity, however, a statistical analysis would be required to determine this interaction.

Chord consonance

Next, the effect of chord consonance on listeners' detection accuracy was investigated. Figure 15 shows subjects' performance with consonant and dissonant chord stimuli (all probe types) in *pure tone* and *piano tone* conditions. S3 performed equally well with consonant and dissonant chords within each condition, better in the *pure tone* condition. The other subjects, with the exception of S4 in the *pure tone* condition, showed a slight trend towards higher accuracy with dissonant chords. This trend is most evident with the professional subjects, S6 and S7, and with piano tone stimuli. An interaction between consonance and sound complexity, or consonance and training would need to be verified with statistical analyses.

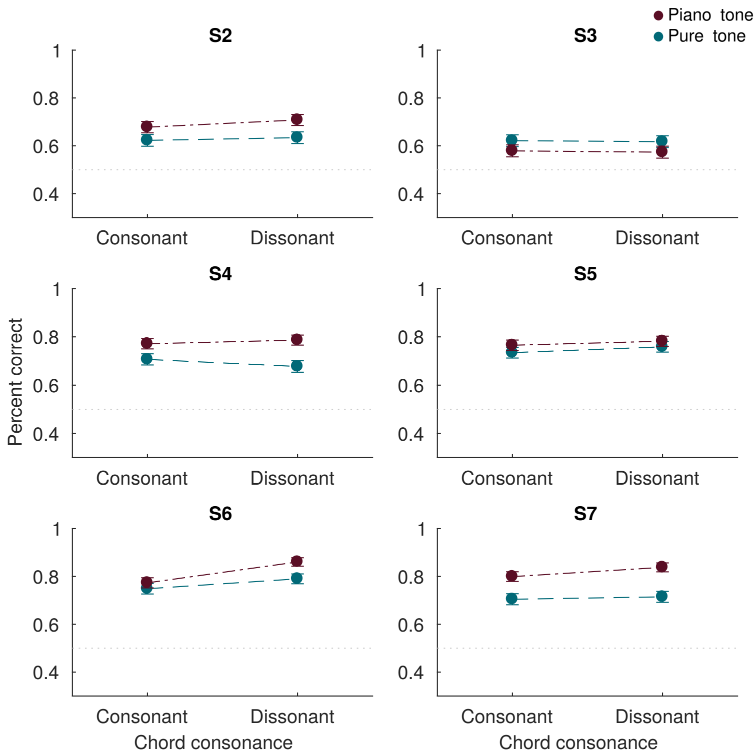

Figure 15: Performance with consonant and dissonant chords in the *piano tone* and *pure tone* conditions with all probe types combined. Error bars show the 95% confidence interval of the normal approximation to the binomial distribution. Grey dotted line represents chance-level performance.

Consonance and position

Figures 16 and 17 show the hit rate (percent correct on signal present/ correct probe trials) of the amateur and naïve listeners, and professional listeners, respectively, in the *piano* and *probe tone* conditions for each probe position. The numbering of probe positions corresponds to ascending pitch. Rows represent different inversions of the chord.

Absolute position

First, the effects of absolute position are described. A common observation across all subjects is that they perform very close to 100% when the probe in position 5, i.e. the top note is probed. Surprisingly, hit rate at this probe position is universally better in the *pure tone* condition, in particular for S3, S4, S5 and S7. Hit rate is also high for the second highest component, probe position 4, although the advantage of pure tones largely disappears. For other probe positions, there is great variance between subjects, as well as within subjects, with different chord structures and tone conditions.

S2 in Figure 16 shows a relatively consistent hit rate pattern in the different probe positions across chord structures and tone conditions. With the exception of probe position 5, hit rate is higher in the *piano tone* condition, being largely below chance for the *pure tone* condition, but performance in both conditions follows similar rising patterns across the different probe positions.

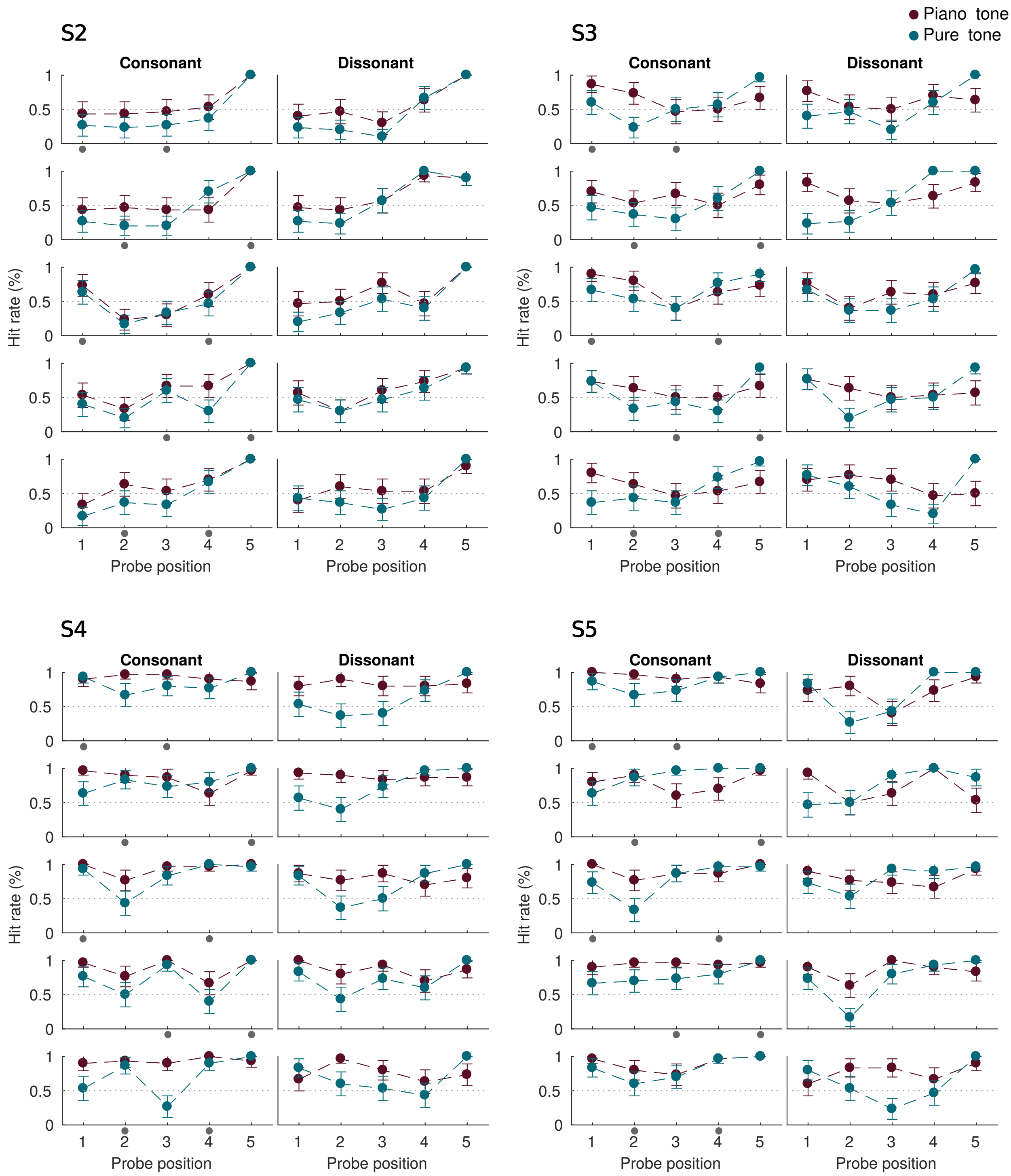

Figure 16: Hit rates of amateur (S2, S4, S5) and naïve (S3) subjects with consonant and dissonant stimuli in the *piano tone* and *pure tone* conditions, at each probe position. Rows represent the different inversions of the consonant and dissonant chords. Grey points on the x-axis represent probes that correspond to the 'root' position of the consonant chords. Error bars show the95% confidence interval of the normal approximation to the binomial distribution. Grey dotted line represents chance-level performance.

S3 shows more variability between the different chords and conditions. Interestingly, the hit rate of S3 is generally higher than that of S2, despite having the lowest performance across all trials. There is something of a U-shaped pattern of performance across probe position for S3, showing highest hit rate with the top and bottom notes of the chord. This trend holds more strongly for the *piano tone* condition because performance at the lower probe positions in some chord shapes (e.g. Diss. 2) is noticeably lower in the pure tone condition.

S4 shows a relatively high and consistent hit rate in the *piano tone* condition across all probe positions. Hit rate in the *pure tone* condition shows great fluctuations between different probe positions which are sometimes also reflected in the *piano tone* condition to a lesser degree (e.g. Cons. 4 and Diss. 4).

S5 appears to show higher and more consistent performance across different probe positions with consonant chords and more fluctuations in hit rate with dissonant chords, however, hit rate varies greatly between different shapes within the same chord categories (e.g. between Cons. 3 and Cons. 4 or between Diss. 3 and Diss. 5).

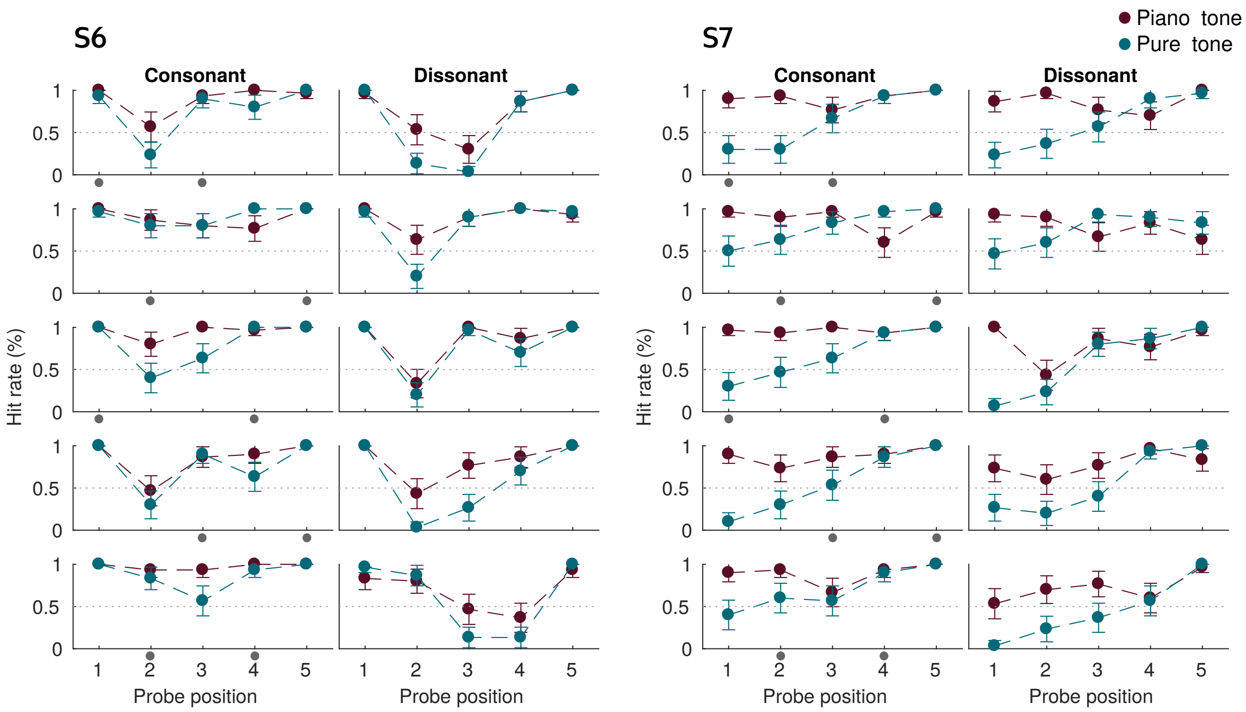

Figure 17: Hit rates of professional subjects with consonant and dissonant stimuli in the *piano tone* and *pure tone* conditions, at each probe position. Rows represent the different inversions of the consonant and dissonant chords. Grey points on the x-axis represent probes that correspond to the 'root' position of the consonant chords. Error bars show the 95% confidence interval of the normal approximation to the binomial distribution. Grey dotted line represents chance-level performance.

In Figure 17, S6 shows equally good performance at probe position 1 as at probe position 5. Hit rates drop at position 2 for most chord shapes, with the notable exception of Diss. 5 where the drop occurs at positions 3 and 4. Generally, when hit rate is high, i.e. markedly above chance, the difference between performance in the *piano tone* and *pure tone* conditions is much smaller than when hit rate is at or below chance.

S7 shows yet a different performance pattern. Hit rates in the *piano* *tone* condition are generally more independent of probe position and is consistently high across all positions in most chord shapes. Performance in the *pure tone* condition, however, increases consistently with probe number, beginning at or below chance at position 1 and gradually approaching performance in the *piano tone* condition.

Functional position

Some of the fluctuations in the consonant chord performance could also be accounted for by the functional position of probes in the major triad. The root notes of the consonant chords are marked on the x-axis of each consonant chord shape with a grey point. According to our hypothesis, these components should be more easily identified than others. In order to distinguish this effect from the effect of absolute position, we focused on the performance fluctuations in middle three probe positions of consonant chords at which performance was generally lower than at positions 1 and 5.

The effect of functional position is particularly strong in S4 and S6. For example, comparing the performance of S4 with Cons. 4 and Cons. 5 reveals a clear reversal of performance patterns in the middle three probe positions, in particular in the pure tone condition (although the same trends are suggested in the piano tone condition). Performance at position 3 is high in Cons. 4 where it coincides with the position of the functional root and low in positions 2 and 4. In Cons. 5, on the other hand, performance is high at positions 2 and 4 where it coincides with the root, and low in position 3. The same pattern is also shown by S6, and to a lesser extent by S2. In other subjects, the effect is either only seen in one tone condition, e.g. S3, or partly masked by the effect of absolute position, e.g. S7.

Certain performance patterns appear across participants in the dissonant chords as well, for example, all subjects show a drop in performance in Diss. 4, at position 2.

Finally, the false alarm rate (percent incorrect on signal absent/ incorrect probe trials) of subjects was explored. The false alarm rate of the naïve and amateur listeners in for each the chords is shown. Figures 18 and 19 show the false alarm rates of naïve and amateur, and professional listeners, respectively, with the incorrect consonant probes for consonant chords (consonant-consonant, column 1), incorrect dissonant probes for consonant chords (consonant-dissonant, column 2) and incorrect dissonant probes for dissonant chords (dissonant-dissonant, column 3). The numbering of probe positions corresponds to ascending pitch. Rows represent different inversions of the chord.

Firstly, Figures 18 and 19 demonstrate that for most subjects, false alarm rate is lowest in the consonant-dissonant trials, followed closely by the dissonant-dissonant trials. The false alarm rate for some probe positions in the consonant-consonant trials, however, is substantially above chance, indicating that subjects consistently misidentified some incorrect probes as having been present in the chord. The figures also show a more similar performance in the *piano* and *pure tone* conditions in the signal absent trials, than in the signal present trials (c.f Figures 16 and 17). The difference between the false alarm rates in *piano* and *pure tone* conditions is greatest in the consonant-consonant trials.

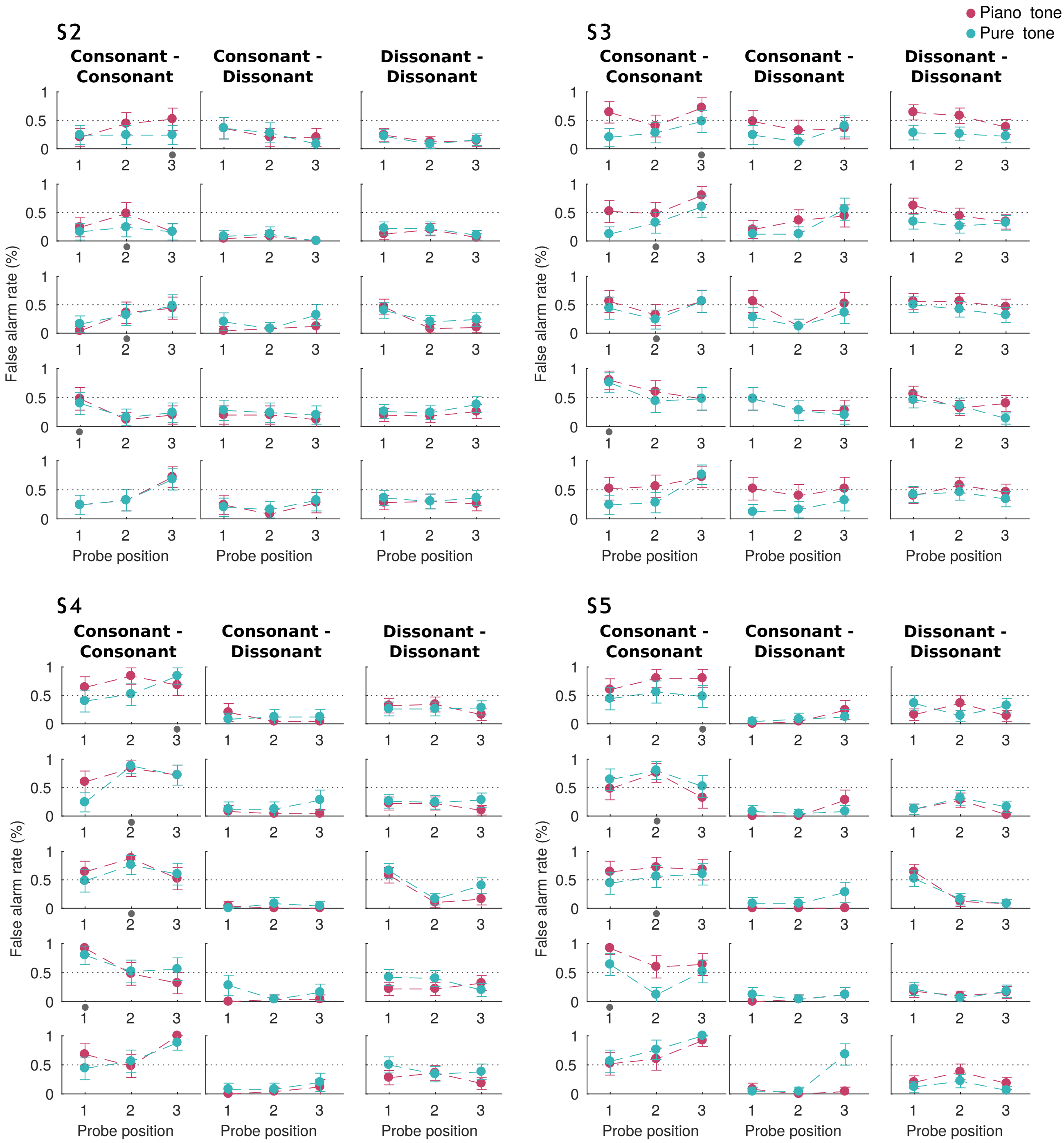

Figure 18: False alarm rates of amateur and naïve subjects with consonant and dissonant chords in *piano tone* and *pure tone* conditions, at each incorrect probe position. Column 1 shows consonant chords with consonant probes (consonant-consonant), column 2 shows consonant chords with dissonant probes (consonant-dissonant) and column 3 shows dissonant chords with dissonant probes (dissonant-dissonant). Rows represent the different inversions of the consonant and dissonant chords. Grey points on the x-axis represent probes that correspond to the root position of the consonant chords. Due to a coding error, probe in position 3 of the Cons. 5 chord was a correct, not incorrect probe, thus this data point actually represents the hit rate. Error bars show the 95% confidence interval of the normal approximation to the binomial distribution. Grey dotted line represents chance-level performance.

Additionally, Figures 18 and 19 show that S2 and S3, the subjects with the least musical experience, had the lowest false alarm rates in the consonant-consonant trials, whereas the professional subjects had the lowest false alarm rates in the consonant-dissonant and dissonant-dissonant trials (with the exception of Cons. 5, position 3 and Diss. 3, position 1 in S6).

In addition to training, subjects' false alarm rates were also affected by the incorrect probes' functional position in the consonant-consonant trials. Probes occupying the root position, marked by a grey dot on the x-axis, were misidentified as having been present more frequently than others, although this effect is not as clear in S2 and S3. The effect is strongest in the false alarm rates of S4, S5, and S6 for chords Cons. 2, Cons. 3 and Cons. 4. The high 'false alarm rate' for all subjects at consonant probe position 3 in Cons. 5 is due to an error in the stimulus code -probe 3 in Cons. 5 was actually a correct probe, thus this data point actually represents the proportion of hits.

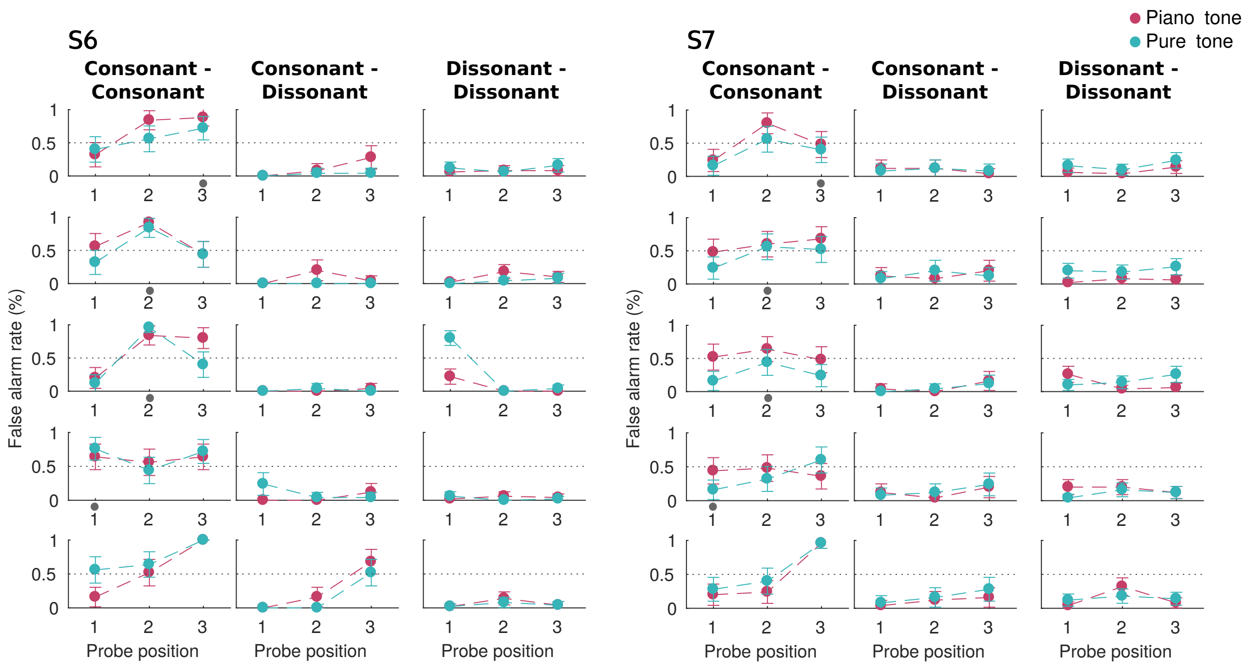

Figure 19: False alarm rates of professional subjects with consonant and dissonant chords in *piano tone* and *pure tone* conditions, at each incorrect probe position. Column 1 shows consonant chords with consonant probes (consonant-consonant), column 2 shows consonant chords with dissonant probes (consonant-dissonant) and column 3 shows dissonant chords with dissonant probes (dissonant-dissonant). Rows represent the different inversions of the consonant and dissonant chords. Grey points on the x-axis represent probes that correspond to the 'root' position of the consonant chords. Due to a coding error, probe in position 3 of the Cons. 5 chord was a correct, not incorrect probe, thus this data point actually represents the hit rate. Error bars show the 95% confidence interval of the normal approximation to the binomial distribution. Grey dotted line represents chance-level performance.

Additionally, S5 and S6 showed a high false alarm rate in the consonant-dissonant trials for probe position 3 in Cons. 5 where the probe was a semitone higher than the root note. S2, S4, S5 and S6 also showed a high false alarm rate in the dissonant-dissonant trials for probe position 1 in Diss. 3 where the probe was a semitone higher than its closest chord component. This component was the same as the starting note for Diss. 1 in that key and thus also corresponded to the root note of the consonant chords in the same key. However, other trials where the incorrect probe had the same relation to the root note of consonant chords in the same key, position 3 for Diss. 1 and position 2 for Diss. 2, did not display high false alarm rates.

Discussion

We investigated listeners' ability to hear the components of musical five-tone chords in a 'yes-no' detection task in which listeners heard a chord, followed by a probe tone which either corresponded to one of the chord components or not. Listeners' performance on this task was affected by the complexity of the tone, the overall consonance of the chord, the consonance of the probe in relation to the chord, the absolute and functional position of the probe in the chord and the listeners' extent of musical training. The effect of each of these factors and their relevance for the harmonicity hypothesis of consonance are discussed below.

Sound complexity

All listeners, except for the naïve subject, performed better when chords were constructed of synthetic piano tones, as opposed to pure tones. The order in which subjects completed the *piano tone* and *pure tone* conditions was counterbalanced but the effect of learning across the four experimental sessions for each individual participant could not be avoided. Since participants (with the exception of S7 in the *pure tone* condition) improved from session 1 of one condition to session 2 of the same condition, it is possible that familiarity with the task and the stimuli improved subjects' ability to correctly identify chord components. It is likely that the naïve subject was more strongly affected by learning effects than other subjects because their performance in the very first experimental session was close to chance. Any advantage to be obtained from the sound complexity in the *piano tone* condition was cancelled by the difficulty of the task in the first session. Other subjects who completed the *piano tone* condition first performed worse in the *pure tone* sessions, demonstrating that the effect of sound complexity overrode the effect of learning.

The finding that chord perception is facilitated when the chord is comprised of naturalistic piano tones is surprising, given that the informational load on the listener is higher in the *piano tone* condition. It suggests that the additional information provided by the harmonic structure is helpful, rather than burdening, in hearing the components of chords. It is possible that the piano tone advantage is simply due to listeners' familiarity with piano-like sounds, or to their unfamiliarity with the highly unnatural pure tones. Professional musicians might have explicit training in identifying the components of piano intervals and chords. At any rate, all listeners are highly likely to have listened to piano music, almost always involving several notes being played at once, which would have implicitly trained them on the experimental task.

On the other hand, it is also possible that the harmonic structure, inherent in any complex tone, piano tones included, provides facilitative cues to the listener which allows them to better parse the composite chord into its constituent components. Harmonic series might, for example, provide richer information about the consonance of the chord which could facilitate the process of deciding whether a certain component was present or not. Indeed, the consonance of the probe in relation to the chord appears to be a major determinant in how frequently incorrect probes are misidentified as having belonged to the chord.

In order to fully distinguish between these two possibilities, an additional experiment should be conducted, using artificially generated complex tones with only a few harmonics which do not sound as familiar as piano tones. If performance with this level of sound complexity resembles that in the *pure tone* condition, it is likely that the piano tone facilitation is caused by the familiarity of the sound. Alternatively, if performance resembles that in the *piano tone* condition, it is more likely that the actual presence of harmonics is the source of facilitation in the *piano tone* condition.

Since most of hearing research, including some research on consonance and dissonance judgments (e.g. Kameoka and Kuriyagawa, 1969; Plomp and Levelt, 1965) and chord perception (e.g. Demany and Ramos, 2005; Schnuerch et al., 2014), uses pure tone stimuli, it is important to note that high performance with complex tones does not necessarily translate equally high (or higher) performance with pure tones. The findings of Demany and Ramos (2005) and Schnuerch et al. (2014) that humans are unable to hear the individual components of chords may be partly due to their use of pure tone stimuli. It is therefore necessary to conduct more comparative studies using both types of stimuli.

Consonance

Subjects were slightly better at identifying chord components with dissonant chords than with consonant chords. Although such a comparison has not yet been studied with chords (Demany and Ramos (2005) used unmusical, inharmonic chords and Schnuerch et al. (2014) used only consonant chords), Seror and Neill (2015) studied the effect of consonance on the identification of interval components. At first sight, our results are not consistent with the findings of Seror and Neill (2015) who found that components of consonant intervals were easier to identify, although even in their study, the advantage was restricted to the bottom component of the interval.

Our results are also inconsistent with the listeners' subjective reports on the task. Several subjects reported to struggle more with hearing the components of dissonant chords although this trend is not borne out by the data. It may be that while listeners subjectively perceived to hear the components of consonant chords better, this also entailed 'hearing' consonant components which were not actually present, and consequently resulted in a higher false alarm rate than with dissonant chords. The false alarm rates for all subjects, except for the naïve subject, were clearly determined by the consonance of the incorrect probe in relation to the chord. The probes associated with dissonant probes were always dissonant, therefore resulting in better overall performance with dissonant chords. In fact, the hit rate for most subjects was higher with consonant than dissonant chords. It may thus be that the difference between our results and those of Seror and Neill (2015) arises from different task designs.

Absolute component position

Different studies have investigated the salience of interval components as a factor of component position. Frequency following responses suggest that the F_0 of the lower component is more strongly reflected in the auditory brainstem than of the higher one (Bidelman and Krishnan, 2009; Lee et al., 2009). Seror and Neill, on the other hand, showed slightly better accuracy and shorter response times in identifying the top component, although the effect was not statistically significant. From the perspective of music, a case could be made for either the bottom note (the bassline) or the top note (the melody line) to be the easiest to detect.

Our results showed that all subjects were nearly perfect at correctly identifying the top note of a chord. There was much more inter-subject variability, however, in how well the remaining notes were identified. While the top note was identified at least as well or better in the *pure tone* condition, at other probe positions, piano tones were heard better. Some subjects followed something of a U-shape, identifying the top and bottom notes better than the intermediate three probe positions (e.g S6). Others followed an ascent in performance from bottom to top, such that lower notes were heard worse than higher notes (S7, *pure tone* condition). Some subjects followed distinctly different patterns in the two tone conditions, demonstrating a high degree of variability in the individual strategies employed by listeners when listening to chords.

The observation that the highest component was best identified in the *pure tone* condition, while lower components were generally best identified in the *piano tone* condition may be explained by the perceived differences in sound quality of the two tone types across the range of pitches used. Pure tones, for example, are perceived very differently depending on their pitch (Hartmann, 2004). High tones are perceived as piercing, while low tones are perceived as dull, accounting for the differences in performance across the range of absolute probe positions. In fact, subjects reported that the perceived loudness of different chord components varied with pitch. Low pure tones were claimed to be less audible than high pure tones. On the other hand, some reported high piano tones to be less audible than low piano tones. To determine the extent to which perceived loudness could account for our findings, it is first important to determine the relative amplitudes of each piano chord component as produced by the synthesiser (the amplitudes of the pure tone stimuli were equal). Secondly, there may be merit in matching perceived loudness of components in different registers in a separate experiment to see if this improves performance in the low pure tone and high piano tone registers.

Functional component position

The functional position of a component is determined by its relation to the root in a triad. Only the consonant chords were derived from the notes of a triad, therefore only the root of the consonant chords could be determined. While the third and the fifth position, corresponding to the middle and the highest note in a triad also have functional roles in music, only the effect of the root position was explored in this thesis. Our results show that both the hit rate and false alarm rate are highest when the component corresponded to the root note of the chord, suggesting that the components occupying a functional root position are most easily heard, even to the point of being heard when they are not actually present.

Additionally, the high false alarm rate with components that fall into the root position implies that the overall sound somehow resembles or reinforces a pitch that is not present. It may be that the percept of a consonant chord is a fusion of its components which reinforces the root frequency, similarly to how consonant intervals appear to fuse into the perception of a single tone more frequently that dissonant ones (Dewitt and Crowder, 1987). However, neither do consonant interval components always fuse, nor were our subjects unable to identify consonant chord components other than the root. To apply the term 'fusion' may be misguided as our findings are better explained by a degree of similarity between the harmonic spectra of the chord and the component. It would be of interest to analyse whether performance for other components could also be explained by their similarity to the chord.To that end, a correlation analysis comparing the harmonic spectra of each chord in the frequency domain with that of each of its associated components should be conducted. The strength of correlation between each chord-probe pair should predict how frequently this component is identified as belonging to the chord. This approach would also allow us to investigate the fluctuations in performance in dissonant chords where no root could be readily identified based on music theory.

Some subjects reported hearing melodic relationships between the chord and the probe which either assisted them (for example by emphasising a movement of a probe from the root to the third position) or hindered them (for example by leading them to respond to a probe as correct simply because it seemed to follow musically from the chord). Such effects also rely on the functional relationships set up by the chord and the probe.

Musical training

The performance of individual subjects reflected their musical experience. Although subjects were categorised into three groups - naïve, amateur and professional musicians - even within these groups, average performance followed the number of years subjects had spent performing music. A regression analysis should be done to more closely determine the relationship between musical training and performance in different conditions and with different stimuli.

The specific type of musical training also appeared to affect how well the professional musicians identified chord components as a function of their absolute position. S6, who was a cellist, having extensive experience in playing in and attending to the lower registers, performed better than other subjects in detecting the bottom note of the chord in both tone conditions. S7, who was a pianist, on the other hand, showed the greatest advantage for hearing the components of piano chords, in relation to pure tone chords. It would be interesting to test a musician experienced in attending to the higher register to see if their performance differs from other subjects, given that every subject performed best at the highest probe position. The expectation that professional musicians would be more strongly affected by the consonance and sound complexity of the chord was not found to be convincing, although slight tendencies could be discerned.

Harmonicity

This experiment was set up under the hypothesis that consonance is determined by how closely the composite harmonic spectrum of a combination of tones resembles that of a single complex tone. Although the theory does not explicitly suggest that the percept of consonant sound combinations would also resemble that of a single tone, it is not completely unfounded to consider the possibility. Both evidence from tonal fusion experiments (Dewitt and Crowder, 1987) and virtual pitch models which account for several psychoacoustic phenomena, such as the ambiguity of pitch of complex tones (Ritsma and Engel, 1964) and the missing fundamental illusion, by evoking a learning-based Gestalt model of auditory processing (Terhardt, 1974a), support the idea that the percept of certain harmonically related sounds may closely resemble that of a single complex sound.

In order to determine whether the resemblance to a single tone translates into the perception of a single tone, we explored whether the components that most closely resemble the overall harmonic structure of a chord have a privileged position in perception. The notes corresponding to the root position of the consonant chords did indeed receive perceptual facilitation due to their functional position, however, this effect was not as strong as that of the absolute position in the case of the top note of the chord. This suggests that while the harmonic relations between chord components play a role in the perception of chord components, perception is even more strongly dictated by the presence or absence of higher components. The facilitation to perception provided by piano tones further implicates the importance of rich harmonic structure in parsing a chord into its constituent components. Nevertheless, in order to determine whether the perception of chords can in any way be likened to the perception of single pitches as a function of their consonance, a detailed analysis of all probe positions and their harmonic similarity to the chord is necessary.

Conclusions

The aim of this thesis was to contribute to our understanding of how simultaneous sounds in a homophonic context are perceived. Previous studies suggest that people are unable to hear out the pitches of individual components of chords, so long as the number of components exceeds three or four (McDermott and Oxenham, 2008). Using musical stimuli, we demonstrated that listeners are clearly capable of identifying the components of chords, although this ability varies from component to component. Some components, namely those that either occupy the top or bottom position of the chord, or correspond to the root note of the chord, are perceived more reliably than those in other positions. This supports the idea that, in addition to masking each other, neighbouring tones can reinforce each other if they share certain harmonic relationships.

We also found that the ability to hear chord components is enhanced when the chord is composed of piano tones, rather than pure tones, and that the position of the component within a chord affects perception differently depending on the type of tones that the chord is composed of. This highlights the importance of using naturalistic complex stimuli in psychoacoustic research, alongside pure tones. Musical training generally improves listeners ability to hear chords as individual components but even subjects with limited training could detect chord components above chance. We also found large individual differences between subjects in how well particular components are perceived.

Although this exploratory examination has been successful in revealing certain tendencies in chord perception, additional statistical and computational analyses should be conducted to verify the results presented thus far, and to extract additional meaning from the complex performance patterns that are not readily apparent on first inspection.

Appendices

Appendix A: Questionnaire

Subject name:
Date:

How many years of formal music training have you had (besides school) – e.g. learning an instrument, singing lessons, choir?

…..

What were you trained in? If you sing in a choir, what voice part do you sing?

…..

How many years of regular experience with performing music do you have (including playing music by yourself or in an informal setting)?

…..

Do you have perfect (absolute) pitch? YES NO

Do you have any form of sound synaesthesia? If yes, what kind?...

Do you have any kind of hearing impairment? If yes, what kind?...

	Very good	Good	Average	Poor	Very poor
How would you rate your natural musical talent (do you think you have a good ear for music)?					
How do your rate your ability to sight read music?					
How do you rate your ability to repeat musical lines after hearing them?					
How do you rate your ability to hear the baseline in a piece?					
How do you rate your ability to hear the individual components in chords?					
	Very easy	Easy	Average	Difficult	Very difficult
How difficult do you find harmonizing with others?					
When you listen to an orchestra or ensemble playing, how difficult do you find hearing out individual instruments?					
Do you find it easy to detect if an instrument or singer is out of tune (i.e. too sharp or too flat)?					

Figure 20: Questionnaire used to evaluate subjects' musical background and skills.

Appendix B: Selection of chords and probes

Chord selection

Compositionally, the same chords and probes were used in the *pure* and *piano tone* conditions, i.e. all stimuli had the exact same pitches. The stimuli varied in sound complexity between the two experiments.

The stimuli consisted of chord shapes with five components and their inversions. Consequently, each chord had five inversions (including the root inversion). Typically, each successive inversion differs from the previous one by only one component in absolute pitch. However, to contain all stimuli within a range that could comfortably be played on a piano, all inversions were transposed down by a multiple of octaves to begin within a pure fifth of the starting note of the root inversion.

In the broadest terms, the stimuli could be divided into consonant and dissonant chords. Consonant chords were derived from the notes of the major triad while dissonant chords were formed by stacking highly dissonant intervals - seconds, tritones, sevenths - on top of one another.

Initially, three consonant and three dissonant root chord shapes were composed. This number of root chord shapes was determined by imposing a set of criteria that each chord had to meet. Firstly, all resulting inversions had to span a range between 2-3 octaves. Secondly, successive notes in the chords should be less than an octave apart. Finally, the top and bottom tones of the chord were not to be a multiple of octaves apart, i.e. the note names of the top and bottom tones were always different. Under these constraints, only three consonant root chord shapes resulted in completely unique inversions. For consistency, the number of dissonant root chord shapes was also limited to three. This resulted in six root chords, with five inversions each (root inversion included), amounting to a total of 30 different chords (15 consonant, 15 dissonant).

This number was later reduced to one consonant and one dissonant root chord shape (total of 10 different chords) due to constraints on testing time. Additionally, using more than one root chord shape in either category would have increased the variance within the dissonant chords more than within the consonant chords, since the consonant root chord shapes were all composed of the notes of a major triad and were therefore all harmonically closely related. Limiting the stimuli to a single root chord shape in either category helped to balance the variance within the two sets of stimuli.

A subset of the unused chord stimuli were used in the practice trials to familiarise participants with the task design without priming them with any of the experimental stimuli.

The absolute pitches of the chords used were not relevant for our experiment and should not affect participants' performance on this particular task. We therefore used five different starting notes for the consonant and dissonant chord shapes (G#2, A#2, C3, D3, E3) and treated the different transpositions of the same chord shape as *functional repetitions*. This created greater perceived variability among stimuli and increased task difficulty while still allowing us to compare a limited set of chord shapes. In total, 50 chords differing in pitch (10 shapes, 5 starting notes) were created.

Probe selection

Each chord shape was originally associated with a set of 11 probe tones. For different transpositions of the same chord shape, resulting from the use of different starting notes, the probes were transposed by a corresponding amount. The five correct probes consisted of the components of the chord. The six incorrect probes were selected from among tones that were not present in the chord but were within the span of the chord. As far as possible, the probes for different inversions employed tones with the same note names, reshuffled to complement the structure of the inversion.

In the case of consonant chords, the incorrect probes were divided into consonant and dissonant probes. The consonant probes consisted of those notes of the major triad which were not present in the chord. For example, the consonant probes for the chord consisting of C3, E3, C4, G4, E5 included G3, E4 and C5. These probes always formed either a minor third, a major third or a perfect fourth with their nearest two chord components. Due to a coding error, one of the incorrect probes (probe 3 for Cons. 5) was a consonant probe. The dissonant probes were selected by altering the consonant components by a few semitones, such that the probe formed a highly dissonant interval with at least one of the two nearest chord components. In the previous example, the dissonant probes were A#5, F#4 and C#5. The dissonant probes formed either minor second, major second or a tritone with at least one of their nearest two chord components.

In the case of dissonant chords, all probes were considered dissonant. They were chosen to produce highly dissonant intervals with at least one of their nearest two chord components. For example, for the dissonant chord consisting of C3, F#3, C4, B4, A#5, the probes were F3, G3, C#4, G#4, E5, and A5. Unfortunately, due to a coding error, only three of the dissonant probes were used in the experiment, however, each dissonant probe was played twice as often to match the number of incorrect consonant chord trials. Additionally, the incorrect probes for Diss. 5 with starting note 5 were inadvertently transposed down by a semitone, resulting in a set of dissonant probes different from all other stimuli. The dissonant probes formed either a minor second, minor sixth or a tritone with at least one of their nearest two chord components.

It is not possible to avoid using any consonant intervals even in the context of dissonant chords and probes, however, care was taken to maximise both the number and dissonance of dissonant intervals.

Appendix C: Generation and presentation of stimuli

Pure tone stimuli were created with the Psychtoolbox (Psychtoolbox 3) extension in MATLAB on a Linux (Ubuntu, 14.04) desktop computer. To create chords, individual sine wave components were first generated. The digital input to the digital-to-analog converter (DAC) only accepts values within a certain range (-1 to 1), values beyond this range are clipped at ceiling, resulting in unpleasant sound artefacts. Therefore, the amplitude of each component was set to 0.2 to ensure that the summed chord does not exceed this maximum range. The five sine waves were then added and the resulting complex was windowed with a symmetric tapered cosine (ratio of cosine-tapered section to the length of the entire window, $\alpha = 0.75$) envelope to prevent the unpleasant clicking noise that results from a sudden sound onset. Pure tone probes consisted of single sine waves with an amplitude of 0.2, windowed with the same tapered cosine envelope.

Piano tone stimuli were created in MuseScore (MuseScore 2.0.3) music notation software. First, the selected chords and probes were manually entered using the music notation features of the programme in quarter notes in 1/4 meter. Each chord and probe was individually exported in .wav format. Due to the reverberation of the piano sounds, the resulting .wav files were much longer than desired and were cropped in MATLAB. To produce a gradual sound onset but retain the characteristic piano onset, an asymmetric tapered cosine envelope (first half of the window (onset) used $\alpha = 0.1$, second half of the window (offset) used $\alpha = 0.75$) was used. The same procedure was followed in creating piano probes. For professional musicians, the amplitude of the probe .wav files was altered in MATLAB to approximate the amplitude of individual the components in the chord, as was done with the pure tone stimuli. This was achieved by dividing the amplitude of the MuseScore output file by the number of chord components. One additional subject, S1, was tested with both the amplitude-corrected and -uncorrected stimuli and the resulting performance is shown in Figure 21.

The loudness of stimulus presentation was controlled through the Psychtoolbox PsychPortAudio sound driver and the Grace m903 amplifier (Grace Design, USA, Figure 22). The sound driver volume settings default to 1 which means that the audio sample is passed to the audio device unmodified in amplitude. Because of the different harmonic and temporal profiles, giving rise to differences in perceived loudness, the same volume settings could not be applied to both the pure and piano tones. First, the volume settings for the pure tone stimuli were selected. The sound driver volume for the pure tone stimuli was fixed to 1. Then, a comfortable listening volume was determined by manipulating the manually adjustable volume settings on the amplifier. The amplifier volume was set to 70 and kept fixed at this value in both experiments.

Figure 22: Grace m903 external sound card and headphone amplifier (Grace Design, USA), front and back.

Next, the sound driver volume for the piano tone experiment was calibrated by comparing the pure tone stimuli at volume 1 with piano tone stimuli at various volume settings, in order to produce similar perceived loudness in the two experiments. As pure tones change in their perceived sound quality across the pitch range used -

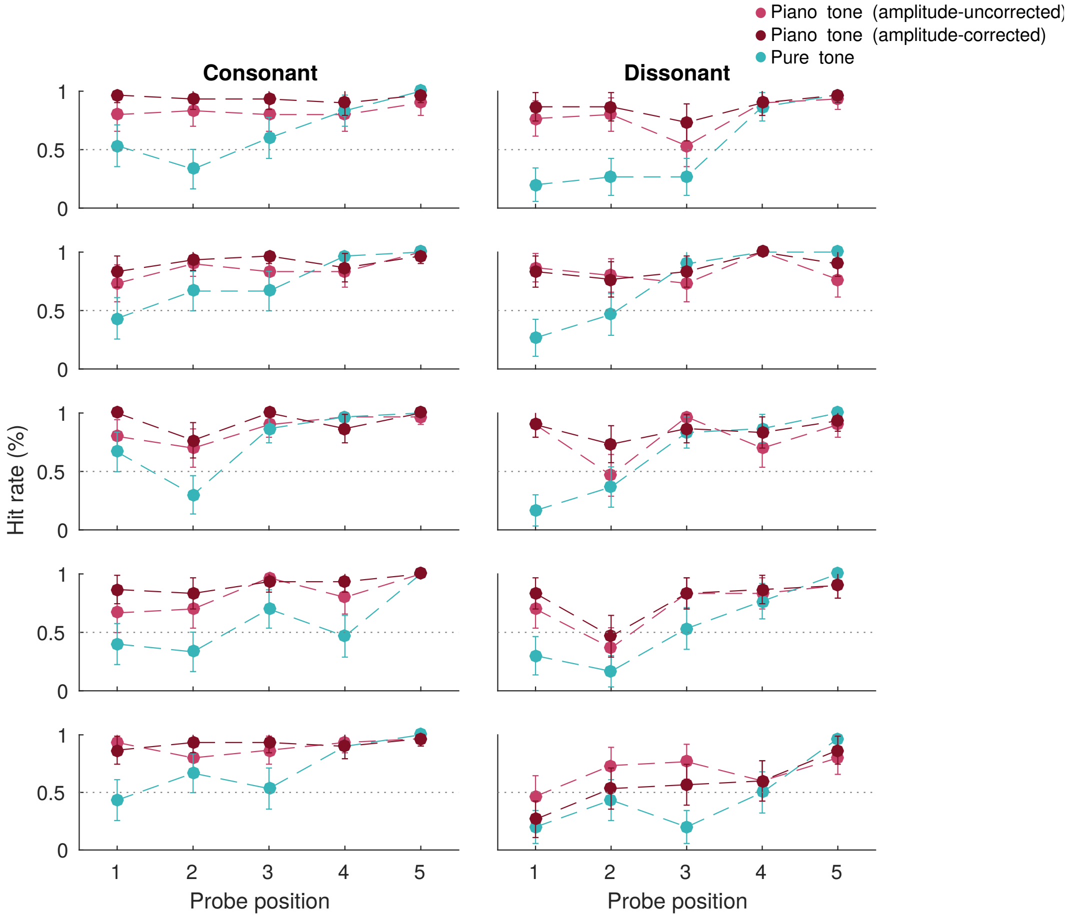

Figure 21: Performance of S1 in the *pure tone*, *piano tone*-amplitude corrected and *piano tone*-amplitude uncorrected conditions, all chords combined. Error bars show the 95% confidence interval of the normal approximation to the binomial distribution.

Figure 23: Ultrasone Signature PRO over-ear headphones with S-Logic PLUS technology (Ultrasone, Germany).

lower tones are dull and quiet while higher tones are loud and piercing (Hartmann, 2004), the piano tone loudness match was selected based on the perceived loudness of mid-range pure tones. Consequently, lower piano tones sounded a bit louder than their pure tone counterparts, while higher piano tones sounded quieter than the corresponding pure tones. The sound driver volume for piano tones was set at 0.4 and the amplifier volume remained at 70.

Gaussian white noise, presented between trials to minimise the impact of successive trials on each other, was generated in MATLAB. Albeit very rare, the standard Gaussian distribution contains values beyond a standard deviation, outside the range accepted by the DAC. In order to prevent unwanted sound artefacts due to clipping of samples above 1 and below -1, the variance of the distribution was compressed and the few remaining values above 1 and below -1 were replaced with 1 and -1, respectively. The sound driver volume for Gaussian noise was set to 0.2 and the amplifier volume remained at 70.

All stimuli were presented via Ultrasone Signature PRO closed-back headphones with S-Logic technology (Ultrasone, Germany, Figure 23).

References

Alvarez, G. A. and P. Cavanagh (2004). "The Capacity of Visual Short-Term Memory is Set Both by Visual Information Load and by Number of Objects". In: *Psychological Science* 15.2, pp. 106–111.

Ballora, M. (2003). *Essentials of Music Technology*. 1st ed. Upper Saddle River, NJ: Prentice Hall.

Beal, A. L. (1981). "Cognitive correlates of musical similarity in chords for musicians and nonmusicians". In: *The Journal of the Acoustical Society of America* 69 (S1), S38–S38.

Bethell-Fox, C. E. and R. N. Shepard (1988). "Mental rotation: Effects of stimulus complexity and familiarity". In: *Journal of Experimental Psychology: Human Perception and Performance* 14.1, pp. 12–23.

Bidelman, G. M. and M. G. Heinz (2011). "Auditory-nerve responses predict pitch attributes related to musical consonance-dissonance for normal and impaired hearinga)". In: *The Journal of the Acoustical Society of America* 130.3, pp. 1488–1502.

Bidelman, G. M. and A. Krishnan (2009). "Neural correlates of consonance, dissonance, and the hierarchy of musical pitch in the human brainstem". In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 29.42, pp. 13165–13171.

Bigand, E., R. Parncutt, and F. Lerdahl (1996). "Perception of musical tension in short chord sequences: The influence of harmonic function, sensory dissonance, horizontal motion, and musical training". In: *Perception & Psychophysics* 58.1, pp. 125–141.

Bitterman, M. E., J. Krauskopf, and J. E. Hochberg (1954). "Threshold for Visual Form: A Diffusion Model". In: *The American Journal of Psychology* 67.2, pp. 205–219.

Briggs, K. D. (2014). *The Language and Materials of Music Third Edition*. Raleigh, NC: Lulu. 581 pp.

Carmichael, L., P. Hogan H., and A. Walter A. (1932). "An experimental study of the effect of language on the reproduction of visually perceived form". In: *Journal of Experimental Psychology* 15.1, pp. 73–86.

Cazden, N. (1980). "The Definition of Consonance and Dissonance". In: *International Review of the Aesthetics and Sociology of Music* 11.2, pp. 123–168.

Colwell, R. (2006). *MENC Handbook of Musical Cognition and Development*. New York, NY: Oxford University Press. 290 pp.

Cousineau, M., J. H. McDermott, and I. Peretz (2012). "The basis of musical consonance as revealed by congenital amusia". In: *Proceedings of the National Academy of Sciences* 109.48, pp. 19858–19863.

Crowder, R. G. (1985). "Perception of the major/minor distinction: II. Experimental investigations". In: *Psychomusicology: A Journal of Research in Music Cognition* 5.1, pp. 3–24.

Darwin, C. (1871). *The Descent of Man, and Selection in Relation to Sex*. London: John Murray.

Demany, L. and C. Ramos (2005). "On the binding of successive sounds: Perceiving shifts in nonperceived pitches". In: *The Journal of the Acoustical Society of America* 117.2, pp. 833–841.

Dewitt, L. A. and R. G. Crowder (1987). "Tonal fusion of consonant musical intervals: The oomph in Stumpf". In: *Perception & Psychophysics* 41.1, pp. 73–84.

Dissanayake, E. (2000). "Antecedents of the temporal arts in early mother-infant interaction". In: *The origins of music.* Ed. by N. Wallin L., B. Merker, and S. Brown. Cambridge, MA: MIT Press.

Dowling, W. J. and D. S. Fujitani (1971). "Contour, Interval, and Pitch Recognition in Memory for Melodies". In: *The Journal of the Acoustical Society of America* 49.2, pp. 524–531.

Edworthy, J. (1985). "Interval and Contour in Melody Processing". In: *Music Perception: An Interdisciplinary Journal* 2.3, pp. 375–388.

Fletcher, H., E. D. Blackham, and R. Stratton (1962). "Quality of Piano Tones". In: *The Journal of the Acoustical Society of America* 34.6, pp. 749–761.

Fujioka, T. et al. (2005). "Automatic Encoding of Polyphonic Melodies in Musicians and Nonmusicians". In: *Journal of Cognitive Neuroscience* 17.10, pp. 1578–1592.

Gibson, D. B. (1986). "The Aural Perception of Nontxaditional Chords in Selected Theoretical Relationships: A Computer-Generated Experiment". In: *Journal of Research in Music Education* 34.1, pp. 5–23.

Hall, D. E. (1974). "Quantitative Evaluation of Musical Scale Tunings". In: *American Journal of Physics* 42.7, pp. 543–552.

Hartmann, W. M. (2004). *Signals, Sound, and Sensation*. New York, NY: Springer Science & Business Media. 680 pp.

Hauser, M. D., N. Chomsky, and W. T. Fitch (2002). "The Faculty of Language: What Is It, Who Has It, and How Did It Evolve?" In: *Science* 298.5598, pp. 1569–1579.

Hauser, M. D. and J. H. McDermott (2003). "The evolution of the music faculty: a comparative perspective". In: *Nature Neuroscience* 6.7, pp. 663–668.

Headphone Measurements (2011). InnerFidelity. URL: `http://www.innerfidelity.com/headphone-measurements` (visited on 10/15/2016).

Heyer, K., B. Ryan, and J. MacDonald (1976). "The relationship between structural stimulus complexity and clarity of iconic storage under a backward masking paradigm". In: *Acta Psychologica* 40.3, pp. 189–198.

Homa, D., B. Haver, and T. Schwartz (1976). "Perceptibility of schematic face stimuli: Evidence for a perceptual Gestalt". In: *Memory & Cognition* 4.2, pp. 176–185.

Hutchinson, W. and L. Knopoff (1978). "The acoustic component of western consonance". In: *Interface* 7.1, pp. 1–29.

Kameoka, A. and M. Kuriyagawa (1969). "Consonance Theory Part I: Consonance of Dyads". In: *The Journal of the Acoustical Society of America* 45.6, pp. 1451–1459.

Kleiner, M. et al. (2007). "What's new in psychtoolbox-3". In: *Perception* 36.14, pp. 1–16.

Lahdelma, I. and T. Eerola (2014). "Single chords convey distinct emotional qualities to both naïve and expert listeners". In: *Psychology of Music*, pp. 37–54.

— (2016). "Mild Dissonance Preferred Over Consonance in Single Chord Perception". In: *i-Perception* 7.3, pp. 1–21.

Lee, K. M. et al. (2009). "Selective Subcortical Enhancement of Musical Intervals in Musicians". In: *The Journal of Neuroscience* 29.18, pp. 5832–5840.
Licklider, J. C. R. (1954). ""Periodicity" Pitch and "Place" Pitch". In: *The Journal of the Acoustical Society of America* 26.5, pp. 945–945.
London, J. (2012). *Hearing in Time: Psychological Aspects of Musical Meter*. New York, NY: Oxford University Press. 252 pp.
Lundin, R. W. (1947). "Toward a Cultural Theory of Consonance". In: *The Journal of Psychology* 23.1, pp. 45–49.
Malmberg, C. F. (1918). "The perception of consonance and dissonance". In: *Psychological Monographs* 25.2, pp. 93–133.
McDermott, J. H. (2009). "What Can Experiments Reveal About the Origins of Music?" In: *Current Directions in Psychological Science* 18.3, pp. 164–168.
— (2013). "Audition". In: *The Oxford Handbook of Cognitive Neurosciences*. Ed. by K. N. Ochsner and S. Kosslyn. 1st ed. Vol. 1. Oxford University Press.
McDermott, J. H., M. V. Keebler, et al. (2010). "Musical intervals and relative pitch: Frequency resolution, not interval resolution, is special". In: *The Journal of the Acoustical Society of America* 128.4, pp. 1943–1951.
McDermott, J. H., A. J. Lehr, and A. J. Oxenham (2010). "Individual Differences Reveal the Basis of Consonance". In: *Current biology* 20.11, pp. 1035–1041.
McDermott, J. H. and A. J. Oxenham (2008). "Music perception, pitch, and the auditory system". In: *Current Opinion in Neurobiology* 18.4, pp. 452–463.
McDermott, J. H., A. F. Schultz, et al. (2016). "Indifference to dissonance in native Amazonians reveals cultural variation in music perception". In: *Nature* 535.7613, pp. 547–550.
McLachlan, N. et al. (2013). "Consonance and pitch". In: *Journal of Experimental Psychology: General* 142.4, pp. 1142–1158.
Miller F., G. (2000). "Evolution of human music through sexual selection". In: *The origins of music.* Ed. by N. Wallin L., B. Merker, and S. Brown. Cambridge, MA: MIT Press.
Oxenham, A. J. (2012). "Pitch perception". In: *The Journal of Neuroscience* 32.39, pp. 13335–13338.
Page, M. F. (2004). "Perfect harmony: A mathematical analysis of four historical tunings". In: *The Journal of the Acoustical Society of America* 116.4, pp. 2416–2426.
Parncutt, R. (2011). "The Tonic as Triad: Key Profiles as Pitch Salience Profiles of Tonic Triads". In: *Music Perception: An Interdisciplinary Journal* 28.4, pp. 333–366.
Parncutt, R. and G. Hair (2011). "Consonance and dissonance in music theory and psychology: Disentangling dissonant dichotomies". In: *Journal of Interdisciplinary Music Studies* 5.2, pp. 119–166.
Plack, C. J. (2010). "Musical Consonance: The Importance of Harmonicity". In: *Current Biology* 20.11, R476–R478.
Plomp, R. and W. J. M. Levelt (1965). "Tonal Consonance and Critical Bandwidth". In: *The Journal of the Acoustical Society of America* 38.4, pp. 548–560.

Rasmussen, M., S. Santurette, and E. MacDonald (2014). "Consonance perception of complex-tone dyads and chords". In: *7th Forum Acusticum*.

Ritsma, R. J. and F. L. Engel (1964). "Pitch of Frequency-Modulated Signals". In: *The Journal of the Acoustical Society of America* 36.9, pp. 1637–1644.

Roederer, J. (2008). *The Physics and Psychophysics of Music: An Introduction*. New York, NY: Springer Science & Business Media. 238 pp.

Schellenberg, E. G. and L. J. Trainor (1996). "Sensory consonance and the perceptual similarity of complex-tone harmonic intervals: Tests of adult and infant listeners". In: *The Journal of the Acoustical Society of America* 100.5, pp. 3321–3328.

Schnuerch, R. et al. (2014). "The Change-Deafness Phenomenon in Harmonic Chords". In: *Swiss Journal of Psychology* 73.3, pp. 143–152.

Schwartz, D. A., C. Q. Howe, and D. Purves (2003). "The statistical structure of human speech sounds predicts musical universals". In: *The Journal of Neuroscience* 23.18, pp. 7160–7168.

Sembos, E. C. (2006). *Principles of Music Theory*. Raleigh, NC: Lulu. 164 pp.

Seror, G. A. and W. T. Neill (2015). "Context Dependent Pitch Perception in Consonant and Dissonant Harmonic Intervals". In: *Music Perception: An Interdisciplinary Journal* 32.5, pp. 460–469.

Terhardt, E. (1974a). "On the Perception of Periodic Sound fluctuations (Roughness)". In: *Acta Acustica united with Acustica* 30.4, pp. 201–213.

Terhardt, E. (1974b). "Pitch, consonance, and harmony". In: *The Journal of the Acoustical Society of America* 55.5, pp. 1061–1069.

Treisman, A. (1985). "Preattentive processing in vision". In: *Computer Vision, Graphics, and Image Processing* 31.2, pp. 156–177.

Zeitlin, L. R. (1964). "Frequency Discrimination of Pure and Complex Tones". In: *The Journal of the Acoustical Society of America* 36.5, pp. 1027–1027.