MASTERARBEIT | MASTER'S THESIS

Titel | Title

Consonant Polyrhythms? Exploring Similarities in Processing of Tone Intervals and Polyrhythms

verfasst von | submitted by

Jonathan David Pietsch BSc

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien | Vienna, 2024

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 066 840
Studienrichtung lt. Studienblatt | Degree programme as it appears on the student record sheet:	Masterstudium Psychologie
Betreut von | Supervisor:	Univ.-Prof. Dr. Ulrich Ansorge

Acknowledgments

I am very grateful for my advisor, Ulrich Ansorge, who enabled and encouraged me to explore this topic and guided me through the development of this paper. Without him, this exploration would still only be in my head. I would also like to thank Martin Lindenbeck and the team at the Acoustic Research Institute at the Austrian Academy of Sciences for helping me with the experimental design and execution, enabling me to carry out my experiment to scientific standards.

Lastly, I want to thank my fiancée Theresa for always being by my side and sacrificing countless hours and nerves. It has been a long journey, and I could not have done it without your support.

Table of Contents

Abstract

The review of Vuust et al. (2022) proposes the predictive coding of music model, which unifies the different musical domains into one framework. Both pitch (Di Stefano et al., 2022) and rhythm processing (Vuust & Witek, 2014) have been studied extensively, we propose a new experimental paradigm based on PCM to find additional similarities of both domains by testing if processing advantages of consonant tone intervals (Basiński et al., 2023) are also affecting polyrhythms with the same ratio. We used a staircase design with a discrimination task, where participants decided which of two similar stimuli are more periodical. Since complexity and tonality of tone intervals highly correlate, the design accommodates measures to identify the source of the effect. While this has only been partly successful, the results show that consonant polyrhythms performed best, which suggests the highest precision in its cognitive model. Additionally, we found that dissonant polyrhythms performed worse than control stimuli. This result cannot be explained by current models, but proposals for mechanisms are made. Due to flaws in the experimental design, we are unable to exclude alternative explanations for these results like familiarity and explanations for the main effect remain inconclusive. Improvements to the experimental design are discussed and we recommend future research in this direction to extend the evidence for PCM and knowledge for both musical domains.

Zusammenfassung

Das „predictive coding of music”-Modell (PCM) von Vuust et al. (2022) vereinigt verschiedene kognitive Verarbeitungsmodelle von allen musikalischen Bereichen wie Töne (Di Stefano et al., 2022) und Rhythmen (Vuust & Witek, 2014) in ein umfassendes Modell der Musikwahrnehmung- und Verarbeitung basierend auf Prinzipien von „predictive coding“ (Rao & Ballard, 1999). Basierend auf PCM schlagen wir ein neues experimentelles Paradigma vor, um Ähnlichkeiten in der Verarbeitung von Bereichen zu finden, die noch nicht untersucht wurden. Es wird experimentell untersucht, ob Verarbeitungsunterschiede, die bisher nur bei Tonintervallen gefunden wurden (Basiński et al., 2023), auch bei Polyrhythmen zu finden sind, die die selben Frequenzverhältnisse haben wie Tonintervalle. Das Staircase-Design beinhaltet eine Diskriminationsaufgabe, wo Teilnehmer:innen zwischen zwei sehr ähnlichen tonalen oder atonalen Polyrhythmen unterschieden sollen. Die Ergebnisse zeigen, dass konsonante Polyrhythmen an genausten unterschieden werden konnten, was auf eine höhere Präzision des mentalen Modells von diesem Rhythmus hindeutet. Des Weiteren wurden bei dissonanten Polyrhythmen die höchsten Unterschiede gemessen, was sich mit keinem derzeit bekannten Modell erklären lässt. Durch Fehler im Versuchsdesign ist es nicht gelungen alternative Erklärungen wie Familiarität eindeutig auszuschließen und den Haupteffekt zu erklären. Für zukünftige Forschung werden Änderungen am Versuchsdesign diskutiert und eine Wiederholung mit verbessertem Design vorgeschlagen, um das Wissen für beide musikalische Bereiche zu erweitern.

1. Introduction

An interesting phenomenon occurs when an old Diesel engine accelerates. At idle, you can hear a rhythmic thump from the cylinders, and it's easy to distinguish the individual strokes. However, as the engine accelerates, the rhythm quickens until the thumps blend, transforming into a continuous rumbling pitch. As the engine revs increase, the pitch rises accordingly. Similarly, in Adam Neely's demonstration (Neely, 2018) accelerating isochronous polyrhythms—multiple rhythms with different meters played simultaneously, typically in periodic patterns—create the sensation of hearing distinct harmonic chords. These chords are based on the ratios between the individual rhythms. While this phenomenon is common in everyday life, the reason why perceptions of sound change so drastically when seemingly only the frequency of beats changes remains elusive.

According to current understanding, the sound of an accelerating engine shifts from one cognitive processing mechanism to another — specifically, from rhythm processing to pitch processing — as demonstrated in the experiment by Ungan and Yagcioglu (2014). The study showed that participants were less precise in differentiating between click trains in the range of 20 to 33 Hz than at frequencies above or below this range, indicating the involvement of two distinct processing mechanisms, as mentioned earlier. Participants reported a period within this range where they were unable to discern whether the stimuli were perceived as a rhythm or a pitch, suggesting an overlap between both mechanisms. The main area where rhythm and pitch are processed is in the auditory cortex, where Brugge et al. (2009) were able to show that different parts of Heschl's gyrus (HG, a part of the auditory cortex) seem to be responsible for distinct rhythm frequencies, which would suggest that the parts of the auditory cortex that process frequencies below the hearing spectrum of 20 Hz simply work slightly different to parts above it. Theoretically, we could describe rhythm processing as the processing of the periodic amplitude envelope of an acoustic input, where the distinct sharp onsets of a beat create rhythm perception, regardless of its underlying frequency content. If rhythm frequency then enters the frequency range of pitch, the amplitude envelope then creates the perception of pitch as the neurons responsible for this frequency are excited. Because of this change in perception, we could assume that both pitch and rhythm processing work differently, but other research independently suggests that both mechanisms process auditory input under the same paradigm of predictive coding (PC) (Kumar & Schönwiesner, 2012; Kumar et al., 2011).

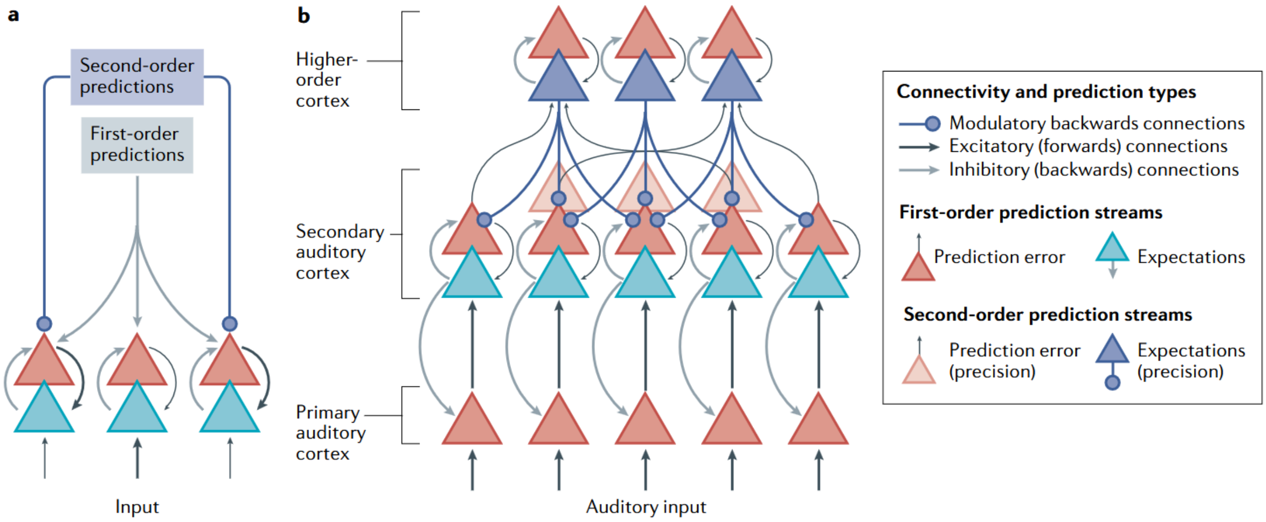
A newer PC-model by Vuust et al. (2022), the predictive coding of music model (PCM), states that all musical domains work with the same underlying model, which processes the acoustic input with the aim to better its internal representation of it by predicting future events of the input. This model enables new kinds of experiments, where it is possible to find similarities between certain attributes of musical domains that are attributed to only this domain. Until now, consonance and dissonance (C/D) are only researched in the domain of pitch processing. We like to challenge this notion and want to find experimentally if C/D is only an attribute of pitch processing or it is a general attribute of PCM, which is currently not yet recognized. If we find evidence for the latter case, we will not only find further support for PCM but also give further insight into PC-processes. As this study is exploratory, we first need to contextualize the current state of rhythm and pitch processing before describing the experimental approach and its results.

2. Theoretical Background

Rhythm and pitch interact most notably during musical listening, which is one of the most complex auditory tasks for the brain. While music encompasses various elements, such as melody, harmony, and rhythm, these components have traditionally been researched mostly independently. They are often regarded as distinct domains, both in music theory and in cognitive processing. But independently from each other, Vuust and Witek (2014) and Kumar and Schönwiesner (2012) among others proposed processing models for each domain that uses PC as a basis, culminating in the PCM by Vuust et al. (2022). The theory states that all musical elements are processed by utilizing auditory input as a basis to predict future events in a hierarchical network. For different timeframes, this internal model of the sensory input postulates hypotheses of what will happen in the future with each element or any combination of them. The aim is to be as precise as possible and to make less mistakes. If a prediction error occurs, this outcome then evokes change other hierarchies as feedback (Figure 1). The model operates using Bayesian inference, with informed priors derived from sensory input and prior experience. This allows the brain to not only process moment-to-moment changes in music but also to perceive the entire composition over a longer timeframe. This capacity to anticipate and integrate musical elements is what often evokes strong emotions and pleasure, for which music is widely renowned.

2.1.1 *Figure 1*

Mechanisms of the predictive coding of music model.



Note. This diagram shows the interactions between the different hierarchical levels. Each level is neuronally connected with feedback- and forward loops, which communicate the predictions made from other levels and gives feedback if these predictions were correct. An incorrect prediction evokes change to reduce prediction errors in the future (From Vuust et al., 2022).

At the moment, PC is a very popular concept to explain the processing of most modalities and was first proposed by Rao and Ballard (1999). Since its publishing, the model is widely used in cognitive research (Friston, 2008; Millidge et al., 2021; Rohrmeier & Koelsch, 2012) and in practically used in neuronal networks (Ciria et al., 2021). Despite its popularity though, many models with PC are criticized due to their difficulty to truly falsify them. See the review of PC for perceptual models by Walsh et al. (2020) for a deeper discussion into this topic.

The research of pitch and rhythm processing sparked a lot of different theories and models about their mechanisms that are difficult to compare but with the PCM in mind, we now look at the current understanding of these musical elements starting with pitch.

2.2 Pitch Processing

One of the most basic elements of music is pitch, which is the perception of a periodic, acoustic waveform. In its simplest form pitch is a sine wave but in practice it contains other frequencies which change the shape of the waveform. The usual components of pitch are: the fundamental frequency (F0), higher harmonic frequencies from the harmonic series (Livshin & Rodet, 2006) and noise. Typically, the lowest frequency is F0 and the frequencies of higher harmonics are derived from it. The content of pitch in its harmonic and non-harmonic content differs and combined with the development of the amplitude envelope over time enables pitch processing to differentiate between the same pitch from different sound sources (Burred et al., 2010). For example, an acoustic guitar usually has a slower attack and pronounced harmonic frequencies than an electric guitar, which makes it acoustically distinct from another.

Processing of acoustic signals starts in the brainstem, where the sensory signal of both cochleae is combined and then sent to the auditory cortex (Oxenham, 2023). In the auditory cortex, neurons are arranged tonotopically, meaning that different frequencies are represented spatially in a way that mirrors the position of excitation along the cochlea in response to acoustic input. The neuronal signal is encoded in the cochlear with two mechanisms at our current understanding: place coding and phase-locking. Place coding relates to the tonotopic arrangement of neurons and phase-locking describes that excited neurons synchronize their output to the frequency of the input. While it is common knowledge that frequencies that can be heard range from 20 to 20.000 Hz, it does not mean that humans are able to perceive pitch at ranges. At the lower end, Brugge et al. (2009) were able to show that click trains lower than the aforementioned threshold are also tonotopically represented in HG, which is a part of the auditory cortex. The upper limit of pitch perception for pure tones is still unclear, as the suggestions for the upper range vary from 1500 up to 10.000 Hz (Verschooten et al., 2019), although the range is typically set at around 5.000 Hz (Attneave & Olson, 1971) for pure tones. More complex tones like chords can be perceived at higher frequencies (Carcagno et al., 2019; Oxenham et al., 2011).

Regarding processing mechanisms, Kumar et al. (2011) were one of the firsts to suggest that pitch processing could be using PC-mechanisms by modeling depth electrode data of HG. They found that the lateral region receives and sends projections comparable to a higher hierarchical level, whereas the medial and middle regions are on a lower level, as described previously. The follow-up paper by Kumar and Schönwiesner (2012) further analyses the results by comparing it to older processing theories, which were mainly focused on bottom-up processes with no feedback from later projections (Cheveigné, 2005). They therefore differed from the results found by Kumar et al. (2011). While this model is only limited to the auditory cortex, other models like the Information Dynamics of Music by Basiński et al. (2023) (originally postulated by Pearce [2005]) were able to expand in it. Before we discuss this model further, we first need to clarify what happens when multiple pitches interact and develop harmony.

2.2.1 Harmony

Harmony in music is the relationship between different pitches in a vertical and horizontal or temporal dimension with tension and release as major qualitative descriptors of different harmonies (Chan et al., 2019). Verticality for harmony describes the different qualitative properties of chords, which differ by stacking pitches with different tone intervals on top of each other. On the other hand, harmony is described horizontally by the qualitative effects occurring from different chord progressions. Due to perceived differences in the quality of different intervals, they are classified historically into either consonant or dissonant intervals (Tenney, 1988). C/D is a fundamental concept in music theory, where consonance is generally associated with pleasantness and stability, while dissonant intervals are perceived as unstable, harsh and with a need to be resolved. The classification of which interval is consonant or dissonant was an extensively debated topic historically, but consonant intervals largely stayed the same over centuries and some argue that the perception of consonance might be hardwired into the brain due to neuronal interactions of the specific harmonic frequencies in the auditory cortex (Tramo et al., 2001).

The perception of C/D intervals differs acoustically as the harmonic interference of consonant intervals is periodic over time, which is often cited as the basis for the association with stability. Physically, the upper pitches of consonant intervals have a small frequency-ratio from the F0 and are part of the harmonic series (Bain, 2003), which describes occurring secondary frequencies when an object vibrates. For example, the tone intervals of a fifth with a ratio of 2:3 is the second harmonic frequency and is perceived as one of the most consonant intervals in music, together with the octave (1:2). While some dissonant intervals are also part of the harmonic series, they are higher up in order and have more numerically complex ratios. At this point, it needs to be noted that we these ratios are derived from the Pythagorean tuning scale, which was superseded by the 12-tone equal temperament scale, which changed the definition of tone interval ratios to be slightly offset from their whole-numbered counterparts in favor of equal distances between all intervals. As such, the consonant intervals we hear today are not as stable as we described earlier but research shows that these differences do not significantly change the perception of intervals (Bailes et al., 2015). While musical experts can discriminate between intervals that have slightly different ratios, non-musicians translate differences to nearby 12-tone intervals and are thus unable to discriminate between just and impure consonant intervals up to a certain degree.

The perceived instability of dissonant intervals is a result of constructive and destructive interference, resulting in different waveforms with irregular patterns and instability of the percept. Helmholtz (1865) termed this quality roughness and it was long the sole explanation for the dissonance perception (McDermott et al., 2010) until newer studies showed discrepancies in this theory. For instance, the results of Itoh et al. (2010) show that the perception of roughness in larger intervals cannot be not explained by roughness alone. In their EEG-experiment, they found that the event-related potential (ERP) associated with roughness only occurred at the three smallest, dissonant intervals, where perception of roughness is also strongest. Other theories regarding C/D differences mainly focus on tonal fusion and familiarity through exposure, which is an integral part of PC-models. Ebeling (2008) describes tonal fusion as a function of periodic firing patterns of neurons, which depend on the tone frequency, resulting in a symmetrical autocorrelation function. Consonant intervals melt into a single percept, where the recipient does not perceive the individual tones, but the sum of its parts. Because of it, this fused percept has the stable properties of consonance, and it often makes it hard for listeners to actively focus attention on a single tone of these intervals.

Neurologically, processing of pitch (and any other acoustic input) starts in the brainstem, where the cochlear-receiving neurons show a more robust salience and strength of signal for consonant intervals, which is in line with the measured behavioral ratings of them (Bidelman, 2013; Bidelman & Krishnan, 2009). These results show also that C/D is not simply a dichotomy, but a range where some intervals are seen as more consonant than others (e.g. an octave versus a third). According to Andermann et al. (2020), this salience translates into processing differences between tone intervals. The observed ERP evoked by dissonant tone intervals showed slight latency differences and a larger onset transient compared to consonant intervals, which indicates longer, more resource demanding processing. Conversely, Crespo-Bojorque et al. (2018) found that consonant chords evoke faster mismatch-negativity potentials (a neuronal activity associated with the processing of unexpected changes) and Park et al. (2011) discovered that consonant chords evoke higher neuronal activity in the right hemisphere after stimuli onset. They interpret the activity as a better integration of consonant chords into a single percept. Similarly to tonal fusion theory, the authors also found that the neuronal activity of processing dissonant chords evoked a less cohesive response, which supports the difficulty of integrating dissonant chords in this theory.

With a similar PC to PCM, Tabas et al. (2019) simulated neuronal processing behavior of tone intervals and came to the same conclusion as mentioned before. They reason that neuronal activity is aided by the period nature of consonant chords as it helps to build up the oscillating activity needed for processing and thus, they decode faster. Another possible explanation for this behavior is proposed by Basiński et al. (2023), who note that consonant chords have lower entropy (less randomness or uncertainty), which eases processing as less information content is needed to achieve an outcome. While they have not yet tested this explanation, it fits into the current understanding C/D processing. For further reading, see the review by Di Stefano et al. (2022).

Expanding to more complex harmonies than discussed before, pitch processing shows a preference towards medium complex harmonies. Li et al. (2021) recorded participants listening to diatonic (simple chords with three tones), chromatic (more complex chords) and atonal chords, which do not follow the rules of western music theory. Their results show that the regions associated with working memory showed the strongest activation with chromatic chords, suggesting a stronger cognitive load compared to diatonic chords. Since diatonic chords are very common in music, participants are very familiar with them which reduces processing load for working memory. Chromatic progressions are more uncommon and often use dissonant intervals, which would evoke more resource intensive activity as prediction errors would be higher within the internal model. As there is no predetermined ruleset for atonal chords, the internal model struggles to make predictions with high confidence leading to an unpleasant listening experience and the inability to predict future musical events. The high activation of chromatic progression is attributed to the change of the internal model as it learns and adapts to harmonies that are unfamiliar but follow a familiar ruleset. This learning experience is theorized to be a main contributor to pleasure and stimulation by music listening (Salimpoor et al., 2015).

2.3 Rhythm Processing

Rhythms are regular acoustic patterns with a distinct duration and follow a meter as its temporal and cognitive framework, which can be divided into multiple levels. A four-bar-rhythm can be mentally divided into four quarter-length or two half-length groups and so on, which changes the perception of the rhythm (London, 2012). The cognitive mechanism of establishing a meter is fundamental to rhythm processing, which is often explained by entrainment, a process of synchronization of the neuronal output with the sensory input. For instance, in the Dynamic Attending Theory (Large & Jones, 1999) the listener's attention is captured by the relative strength of regular metric accents within the rhythmical hierarchy and the internal meter forms by the relationship of both top-down and bottom-up processing. The occurrence of meter and the perception of accents thus happens simultaneously by entrainment with the focus of attention being influenced by accents. The listener is able to predict future events due to the stability of the entrained meter over a short period of time and the general familiarity with rhythmic patterns due to exposure in music (Hallam et al., 2012). While predictions of beats and future musical developments based on entrainment over short periods of time are possible, the limitation of the theory to a singular meter makes it relatively inflexible towards more complex rhythms, compared to newer theories. Because of these limitations, newer models often incorporate PC since most rhythms in music are not simply periodic/isochronous but more complex and often have slight deviations like syncopation and groove (Vuust & Witek, 2014).

The model by Vuust et al. (2018), based on the initial theory by Vuust and Witek (2014), proposes that the internal meter is established by processes based on PC. As the acoustic input is processed, predictions are made for the future occurrence of beats and accents and prediction errors change the internal meter moving forward. More complex or irregular rhythms thus enable the formulation of a more complex meter and flexible predictions which integrate these irregularities into the internal representation of the rhythm. The model has an emphasis on precision or confidence of predictions, which describes the uncertainty towards new predictions and is inversely correlated with prediction error. The higher the precision of a prediction, the higher the prediction error, as the perceived input was more unexpected than if the precision were lower. This relationship can be measured by the participant's behavior, for instance when tapping with a finger to a stimulus, which is a common research paradigm as rhythm perception is often accompanied by sensorimotor synchronization (Stupacher, 2019).

A different PC-model proposed by van der Weij et al. (2017) has a different angle towards the generation of the internal meter. Due to the familiarity of common rhythms within a culture, the internal model proposed by van der Weij et al. generates the meter by aligning the rhythmical input into a mental grid based in prior experience and changes depending on the amount of prediction errors. Their study is based on a simulation, where different models were trained with typical German and Chinese rhythms used in their respective music. After training, their predictive performance was then tested with each data set. The comparison showed that the model trained with the same data set performed better in classifying the rhythms and needed a lower information content within the model for this performance, meaning that it was able to predict a familiar rhythm better with less information necessary. While this probabilistic simulation does fall short of being an exhaustive model of rhythm perception, it emphasizes the importance of an informed prior within the hierarchical model to be successful. This was not expanded upon in the previously discussed model by Vuust et al. (2018).

Similar to neurological evidence of different pitch processing models, rhythm processing has a frequency-following response (FFR) on the neuronal level with distinct bursts of activity around each beat (Brugge et al., 2009). In the study of Nourski et al. (2013), this behavior was most prominent at 50 Hz but occurred up to 200 Hz with limited amplitude in HG, which is well within the established hearing range for pitch, although this behavior was only measured with isochronous beats. More complex rhythms show a more complex neuronal behavior, as explored by Nozaradan et al. (2018). They found that the development of meter in HG can coincide with a more abstract FFR in higher cortical levels, which follow meter-driven frequencies instead of beat-frequencies and is formed by interactions in lower cortical areas, which have a response closer to beat frequencies. The outputs of early processing are then projected to other areas like the cerebellum, which is mainly associated with movement control which controls other processing related behavior. For instance, sensorimotor synchronization often coincides with rhythm processing like finger tapping or other forms of expression and is also part of the experience of flow, which can happen when listening to music (Stupacher, 2019). Rufener et al. (2020) were also able to causally link the cerebellum with the ability to predict regular auditory events by temporary impairing the cerebellum, although this has not been yet replicated with musical stimuli. These cortical behaviors do not contradict current PC-models for rhythm processing, although some rightfully point out that PC is very hard to prove (Walsh et al., 2020).

2.3.1 Polyrhythms and Syncopation

Rhythmic complexity is a strong moderator within most rhythm processing models of both behavioral and neuronal output. This not only entails the complexity or frequency of beats like a classic four-on-the-floor beat versus more complex patterns used in jazz or other progressive musical genres, but also syncopations/groove (slight but systematic timing variations of specific beats against the grid) and specifically polyrhythms. Polyrhythms are rhythms with no or more than one formal meter present in the acoustic signal and the internal meter needs to be inferred by the internal model (Vuust & Witek, 2014). Lumaca et al. (2019) studied the effects of rhythm complexity and found that timing violations lead to higher MMN for larger deviants, while smaller deviations on yielded MMN with a smaller amplitude suggesting a qualitative difference between both stimuli. The authors suggest that smaller deviations do not conflict with the internal meter, as the mental beat pattern stays intact. But perception of larger timing deviations result in a shift of cognitive resources towards the event. This shift in cognitive resources could be because the prediction errors for these deviations grow larger and thus the model changes, which demands the aforementioned resources. This explanation is not explicitly stated by the authors and thus needs to be verified.

While syncopations challenge the internal model due to timing differences in the range of milliseconds, polyrhythms are perceptually challenging as they can challenge the internal model by having multiple rhythmical frameworks at the same time, as explained in detail by Vuust and Witek (2014). As such, polyrhythms are bistable percepts, which perception can change depending on the attended framework. This reference is used to integrate the other framework into one percept, but when a shift the meter changes with it and as such the whole mental model. This process demands more cognitive resources than simpler rhythms, especially if you consider that rhythm gives other musical domains context. If the perceptual context of a polyrhythm shifts, then the perception of other domains can change with it. In popular music, polyrhythms are often expressed as two isochronous rhythms, where one has three beats in the same timeframe as the other has four. In this example, the polyrhythm has a ratio of 3:4 (in music theory often referred to as "Three-against-four") and is often used in popular music, while more numerically complex polyrhythms are only found in more experimental music.

In their study of processing of polyrhythms, Møller et al. (2021) found that listeners tapped their fingers more often two the binary divisible rhythm (e.g., the "four" in 3:4) and were subdividing each polyrhythm into smaller chunks with equal time, which appeared to be much more challenging with more complex polyrhythms like 4:5 or 5:6. This shows a preference for simplicity, likely due to the higher cognitive demand of mentally representing more complex polyrhythms in the mental model through entrainment, which can continue even after the stimulus stopped playing. Stupacher et al. (2017) showed that a particular internal model of a stimulus persists over time, as participants were able to correctly determine the timing of a previous experienced polyrhythm after a slight pause. The neuronal oscillations found during listening continued through the pause, indicating that the internal model was keeping track of the rhythm after it stopped. While there was no difference in the perceived entrainment process of nonmusicians and musicians in this study, the latter showed stronger neuronal oscillations and were more precise with their predictions. This indicates that musicians in this study had a more precise internal model of the stimuli, which lead to more confidence in the prediction of the beat onset. Pallesen et al. (2015) found a similar pattern, when musicians and nonmusicians were listening to different chords. Like the previous study, they discovered that musicians had oscillations with larger amplitudes in the same frequency band, while also having stronger phase locking activity depending on the chord. The frequency of phase locking in this case does not refer to the F0 of the beat, but to the frequency of the amplitude envelope, which changes periodically due to harmonic interference. Thus, it is important to point out that both results are not directly comparable as the overlapping findings are the result of different domains. Nevertheless, this shows that in both domains musicians show more precision and performance in experiments than nonmusicians. This pattern holds true in most other musical domains as well (Heggli et al., 2019; Lahdelma & Eerola, 2020; Strait et al., 2010; Zhao et al., 2017).

As we have seen, both pitch and rhythm processing both have their idiosyncrasies and are precisely tuned to their respective auditory domain, but similarly to the origin of PC in the visual domain by Rao and Ballard (1999), the underlying process can be very similar. With PCM, we now have a framework that emphasizes equivalences over different domains, and we will now discuss the possibility that polyrhythms can have processing idiosyncrasies of tone intervals.

3. Aim of the Experiment

The PCM (Vuust et al., 2022) is a framework for a more holistic approach to cognitive research of music, as it shows that different musical domains share an equivalent process for its perception. This model does not differ much from previous models in each domain (for instance, Virtala & Tervaniemi, 2017; Vuust and Witek, 2014) and thus does not invalidate previous research but can be used to gain more interdisciplinary insight into cognitive modeling and music. The current experiment aims to do exactly that with tone equivalence: As we discussed, both polyrhythms and harmonics can be described in relative ratios (a ratio of 3:4 can be both a perfect fourth and a common ”three-against-four”-polyrhythm) and can be perceived interchangeably if accelerated or decelerated (Neely, 2018). One of these proposed equivalences could be that the differences found in C/D in pitch processing, like the apparent processing advantage of consonant intervals (Basiński et al., 2023; Tabas et al., 2019), are also present in rhythm processing. Until now, there was no practical use for the resulting dissonant polyrhythms as they can be very complex and it could be the case that it was missed, because of it even though they exist. Most practical applications of polyrhythms in music are equal to consonant intervals (for instance, 2:3 equals a perfect fifth), as they have simpler ratios that can be feasibly reproduced by humans. With no practical need for and insight into these complex polyrhythms without a framework, it could be the case that this part of rhythm processing was overlooked in previous theories. As such, we will utilize the perceived connections of rhythm and pitch, to see, if polyrhythms can be indeed consonant or if their overwhelming presence in music compared to dissonant polyrhythms is simply due to their (much) lower complexity.

4. Method

4.1 Study Design

The design uses a performance task similar to the experiment used by Ungan and Yagcioglu (2014), where they used an adaptive testing paradigm with a discrimination task. Participants had to decide if they perceived a change in a click train, which had two components with a similar inter-click-interval with a hearing task. Depending on their response, the components will become more or less similar to each other depending on the success of preceding trials resulting in thresholds of just-noticeable differences (TJND) for the respective frequencies Ungan and Yagcioglu were looking at. Instead of isochronous click trains, the stimuli in our design are polyrhythms with ratios of tone intervals (tonal) or with no tonal relationship (atonal) as control variables. The experiment is designed as within-subject to exclude interpersonal differences and to make this the study more robust (Oberfeld & Franke, 2013). As previously mentioned, musical training can have a strong modulating effect on musical studies. Thus, a questionnaire to establish the musical training of each participant was used to control possible confounding effects.

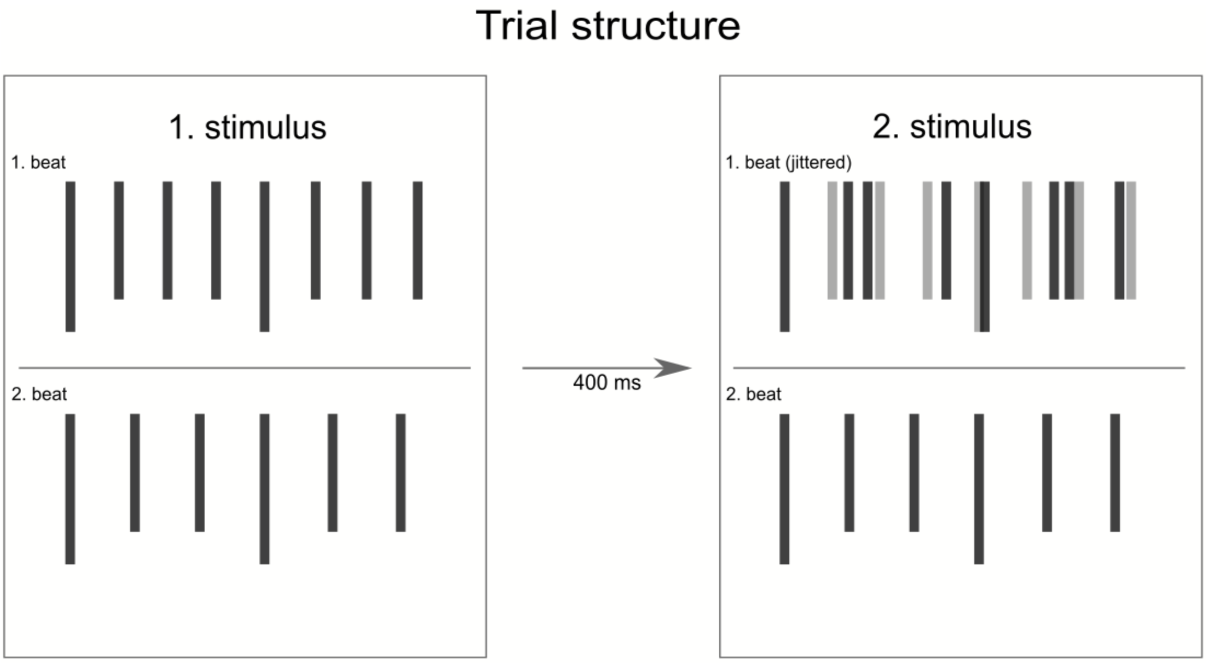
4.2 Procedure

The whole experiment lasts for about one hour including the short questionnaire about previous musical training and experience. Participants were then instructed to sit down in a sound isolated chamber with a computer and studio headphones inside. Their task for each trial was to decide which of the two presented stimuli is the more regular one by pressing keyboard buttons. After receiving short visual feedback if the answer was correct, participants continued to the next trial by another button press. This ensures that participants have the option to take short breaks, if necessary, as it was a demanding task.

The main experiment runs in a custom program ("PolyRhy", version 1.05) based on the ExpSuite-Framework (Majdak & Mihocic, 2020) at a sampling rate of 48 kHz. Participants wore headphones and the volume was adjusted to a comfortable level. Each experiment consisted of one staircase for each ratio. Each trial consisted of a regular and a jittered stimulus in randomized order with a 400 ms break in between with the decision-task directly after. During the trial, each stimulus was represented visually by a numbered box, which lit up green when it was playing. The trials of each staircase were interleaved, and their order randomized to further decrease training effects from previous trials.

4.2.1 Figure 2

Example of a trial



Note. The first stimulus is an isochronous polyrhythm in the ratio of 3:4 (a different tonal low-complexity polyrhythm). After a pause of 400 milliseconds, the participants hear the second stimulus with the same ratio, but one rhythm is jittered. After the first beat, each following beat has a slight randomized offset to its regular timing. The extent of this offset is determined by the JF. (The longer bar represents the start of a new bar but is acoustically identical with the other beats.)

If participants answer correctly two consecutive times within a staircase, the JF is then decreased for the next trial of the staircase but if they guess incorrectly the JF is increased the next trial. After 12 turnarounds, where the JF increased after it decreased before and vice versa, the staircase is considered finished. A staircase would also end if a participant reached JF-values of 0 or below, as a JF of 0 means no difference between the stimuli in a trial and negative values would inverse the experiment.

Initial power analysis was conducted with the WebPower-package in R (Zhang & Yuan, 2018) to determine the minimum sample size for this experiment. Results indicated the required sample size to achieve 80 % power for detecting an expected medium effect at an $\alpha = .05$ was $N = 40$ for a repeated measures ANOVA. As the effect size could only be assumed, we conducted an additional power analysis by bootstrapping the first four results similar to Yuan and Hayashi (2003), which suggest that $N = 20$ should suffice to reach the established target of 80% power and was thus chosen to be our target sample size.

4.2.2 Stimuli

Each stimulus consists of two rhythms played simultaneously in mono resulting in a polyrhythm. In total, four different ratios are used: a low-complexity tonal, a high-complexity tonal and two atonal polyrhythms which also have low- and high complexity (Table 1). Both atonal polyrhythms are used as control for both complexity and tonality, as they are mathematically correlated (a low-complex ratio is considered consonant, while all high-complex ratios are dissonant) and both dimensions cannot be physically separated on the stimuli level. But with two control stimuli, the analysis of both dimensions is possible by pooling and comparing the pools against each other and thus separating both dimensions, even though tonal and atonal polyrhythms might not be perfectly comparable due to the preference for binary polyrhythms (Møller et al., 2021) and atonal polyrhythms have to be nonbinary to not be accidentally tonal, especially for a low-complexity one.

The stimuli are generated at the start of each trial with a gaussian-modulated sinusoidal pulse using the *gauspuls*-function in MATLAB with F0 = 1000 Hz as beat sample. Other possible shapes like sinewave, white noise or square wave were considered, but this shape was chosen for its distinct “click”, which aids participants with the precise onset of each sample. The sample has a length of 3 ms and the spaces in between them were filled with silence. If two samples overlap, they are exchanged for a sample with an amplitude of -9 dBFS to simulate the doubling of acoustic energy when two beats intersect (the regular sample has an amplitude of -12 dBFS). This also avoids acoustic artifacts, which occurred when two samples overlapped. During pilot testing, these artifacts were used to differentiate between rhythms and thus skewed the results. The tempo of the stimuli varies uniformly within a range of 110 to 130 BPM, and they have an average length of 5 seconds.

Table 1

Table of ratios used in the experiment.

	Tonal	Atonal
Low complexity	2:3 (LC)	3:5 (LA)
High complexity	16:9 (HD)	17:7 (HA)

Note. The low-complexity, tonal ratio LC has the ratio to a fifth and the high-complexity, tonal ratio HD is equivalent to a minor seventh. The atonal ratios are not associated to tone intervals but are numerically close to the tonal polyrhythms.

4.2.3 Measures: TJND and Musicality

TJND is defined as the threshold of JF-values, where participants are barely able to recognize a difference between the two stimuli in a trial. High final TJND in this paradigm indicate that the participant’s internal model was not precise enough to notice more subtle differences and was completing the task with low confidence, while low TJND can be a result of comparative differences in the stimuli in the complexity (Crespo-Bojorque et al., 2018; Møller et al., 2021) or tonal equivalence (Basiński et al., 2023; Tabas et al., 2019). A familiarity effect based on exposure in music can only be attributed to LC, as this is the only polyrhythm we use that is used in popular music more often. Unfortunately, we cannot control this factor directly due to numerical constraints, but it should only enhance differences and can be indirectly controlled by pooling the stimuli. To control musical training, a questionnaire was used to collect information about the musical training and ability of each participant. After a brief demographic section, it asked six questions in German. See the Appendix for the full questionnaire.

4.2.4 Research Question 1

The experiment is designed to enable the separation of complexity and tonality of rhythms, when stimuli are pooled according to their categorization within each dimension. If we find a similar processing advantage of C/D due to tonal equivalence (Basiński et al., 2023; Tabas et al., 2019) in rhythm processing, then we should find differences in the tonal stimuli itself and compared to atonal stimuli. Since western audiences are very familiar with tone intervals (Bailes et al., 2015), this should lead to the performance benefit to tonal stimuli. If differences occur mainly due to different complexity levels of stimuli, then low and high complex stimuli should have similar outcomes and thus dismiss the hypotheses of tonal equivalence. With these predictions in mind, the hypotheses are the following:

Hypotheses

H1a (Tonality I)	The TJND of the tonal pool (LC, HD) is lower than the mean of the atonal pool (LA, HA).
H1b (Complexity)	The TJND of the low complexity pool (LC, LA) is lower than the mean of the high complexity pool (HD, HA).
H1c (Tonality II)	The TJND of LC is significantly lower than the TJND of HD.
H1d (Control Group)	The TJND of LA and HA are significantly different.

4.2.5 Research Question 2

Musical training not only enhances the skills in an instrument but also cognitive processing of music, like the precision of both rhythm and harmonic perception. As such, we predict that participants with more musical training, especially rhythmical training, will have lower TJND over tonal and atonal stimuli. Participants are considered musically trained, when they had training in an instrument or in general for at least five years. A participant was considered a professional musician that either has a musical degree or has played music in a professional manner.

Hypotheses

H1e (Musical Training)	Untrained participants have significantly higher TJND then trained or professional participants in the tonal pool (LC, HD).
H1f (Musical Training Control)	Untrained participants have significantly higher TJND then trained or professional participants in the atonal pool (LA, HA).

4.3 Sample

All 19 participants (13 female, $\bar{x} = 22.5 \pm 3.01$ years) were sampled with the LABS panel of the University of Vienna, a university-wide panel for sourcing participants, where they are rewarded with credits, that are a condition for completing particular seminars. After completion all participants were rewarded five LABS-Credits for their attendance. All of them were students in different fields and half of them were at some point in their lives musically active and played an instrument but most of them on an amateur level. None of them had any musical degree, although one participant played drums on an advanced level in the past. 10 participants were categorized as untrained, while 9 were trained based on their survey results. Due to the non-selective sampling process, no professional musician attended the experiment. The following analysis is adjusted to this circumstance.

5. Results

5.1 Survey and Experimental Results

All participants were tested with no dropouts during testing (Table 1). The trained und untrained participants did not differ significantly in age, t_{age} (16.25) = 0.2, p = .806, and their assessment in their ability to follow a rhythm, t_{RF} (16.88) = -1.66, p = .116. As expected, the trained group was significantly more active in their past, U_{active} = 9.5, p >.001, r = -.75, and played an instrument for longer, U_{IY} = 5.5, p >.001, r = -.79. Some of the trained participants have actively played their instruments for multiple years but noted that they have not played for some time at time of testing. They were still included in the trained group as the processing abilities gained from training should not change drastically over time (Hyde et al., 2009).

All participants successfully completed the experiment with 52 trials on average (SD = 15.22) per staircase. In two instances participants were able to hit JFs below 0, which triggered an early termination of the staircase to ensure data quality. Due to the number of trials to achieve JFs below 0, it is very unlikely that these outcomes were achieved by chance. These staircases were included by averaging the JF of the last 6 trials before termination as an estimate of the TJND for these staircases. For the rest, the JF of the last 6 turnarounds overall were averaged to estimate the true TJND (Figure 3A, Table 3) after converting JF into TJND by dividing each inter-beat intervals of every trial with the corresponding JF. This was necessary to normalize against the tempo randomization, since JF was only relative. The TJND of LC were the lowest with $\bar{x}_{LC}$ = 11.66ms against the fully isochronous stimulus, followed by $\bar{x}_{LA}$ = 27.11ms and $\bar{x}_{HA}$ = 24.53ms. HD, the high-complex dissonant stimulus was the highest with $\bar{x}_{HD}$ = 41.65ms.

5.1.1 Table 1

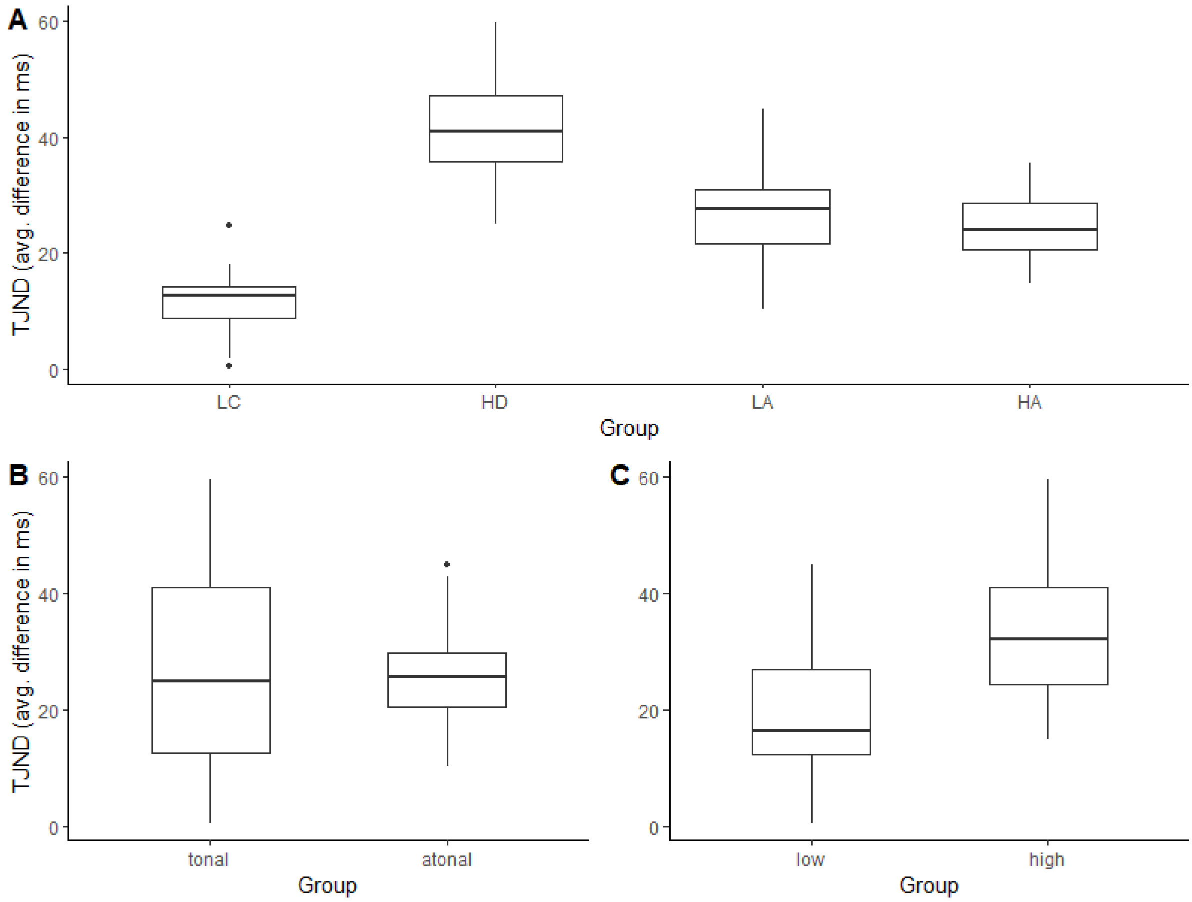
Survey results of all participants.

	$\bar{x}_{all}$	$\bar{x}_{untrained}$	$\bar{x}_{trained}$	*t*	*DF*	*p*
Sample size	19	10	9	-	-	-
Female	13	6	7	-	-	-
Rhythm. Instrument	0.05	0	0.11	-	-	-
Musical degree	0	0	0	-	-	-
Age	22.5 (± 3.01)	22.7 (± 3.02)	22.3 (± 3.35)	0.2	16.25	.806
Musically Active	0.47	0.1	0.89	9.5 (*U*)	-	<.001***
Instrument (Years)	4.21 (± 5.5)	0.2 (± 0.63)	8.67 (± 5.02)	5.5 (*U*)	-	<.001***
Follow rhythm	4.89 (± 1.44)	4.4 (± 1.5)	5.4 (± 1.24)	-1.66	16.88	.116
Musical education	4.3 (± 5.17)	0.8 (± 1.93)	8.2 (± 4.84)	4.5 (*U*)	-	<.001***

Note. The most played instruments were piano and flute, only one participant played a rhythmical instrument. Musical education measured in years was also significantly different between both groups, but quality of education could only be ascertained anecdotally which weakens the validity of this measure. For “musically active”, “instrument (years)” and “musical education” Mann-Whitney-*U* results are reported since they are categorical variables.

5.1.2 Figure 3

Boxplots of stimuli groups and pooled tonality and complexity groups.



Note. The boxplots are corrected for the within-subject design as described by Cousineau (2005). **A** shows the means of all stimuli groups, which are visually different with LC being the lowest and HD being the highest. LA and HA look similarly positioned. **B** and **C** show the pooled groups as referenced in Table 2. The boxplots of the tonality pools have a similar mean, but the tonal group has a much larger variance, while the means of the complexity pools deviate with similar variance.

5.2 Statistical Analysis

To test, if the visual differences are statistically significant, all groups were modelled in a repeated measures ANOVA after testing the assumption of sphericity which was closely nonsignificant with $W = .55$, $p = .079$. The model confirmed the visual findings that the stimuli group has a significant effect on the TJND, F_{rm} (3, 54) = 36.53, p <.001, η^2 = .67, which is a very large effect size (η^2 = .14 typically indicates a large effect size). As the sphericity assumption was close to significance, which could lead to wrong results, we also conducted a robust repeated measures ANOVA without the condition of sphericity to be sure. This ANOVA also had a significant result, F_{rrm} (2.35, 28.2) = 36.63, p >.001, so we continued by calculating orthogonal contrasts with the data of the repeated measures ANOVA (Table 2).

5.2.1 Table 2

Results of orthogonal contrasts and further post-hoc tests.

		DF	*t*	*p*	*d*
rm ANOVA	Overall	3, 54	(*F*) 36.53	<.001 ***	(η^2) .67
contrasts	HD - LC	54	10.71	<.001 ***	-2.52
	LA - LC	54	5.52	<.001 ***	-1.29
	HA - LC	54	4.60	<.001 ***	-1.43
	LA - HD	54	-5.19	<.001 ***	0.87
	HA - HD	54	-6.11	<.001 ***	1.73
	HA - LA	54	-0.92	.439	0.18
post-hoc	Tonal - Atonal	37	0.26	.798	0.04
	Low - High	37	-7.78	<.001 ***	-0.65
	Tonality/Training	35.91	-0.27	.788	-0.08
	Complexity/Training	32.75	0.55	.586	0.18
	TJND/Active	16.46	-0.11	.917	-0.05
	TJND/Gender	13.23	0.16	.874	0.08

Note. Table 4 shows the results of the contrasts and further post-hoc analyses, which were conducted with paired T-tests. Post-hoc tests were: Pooled TJND by tonality and complexity, as well as tonality, complexity, musical activity, and gender grouped by musical training.

Contrast analysis revealed that all groups differ significantly from each other, except $t_{LA\text{-}HA}$ (54) = .47, p =.711, which is the difference between the atonal stimuli (Table 4). The effect sizes of the significant contrasts are all exceptionally large, with most of them with a Cohen's d > 1, indicating very large group differences (typically values of > .8 are considered large). To entangle tonality and complexity, further post-hoc tests were conducted with pooled groups by splitting the groups in equal sized two pools (tonality: LC/HD vs. LA/HA, complexity: LC/LA vs. HD/HA; see Figure 3B and 3C for their boxplots) and then testing for differences between them. The pairwise t-tests for the complexity-pools show significant differences, t_{comp} (37) = -7.78, p = <.001, while the differences in the tonality-pool are nonsignificant, t_{ton} (37) = 0.26, p = .798, which is surprising as the non-pooled means for tonal stimuli were significantly different.

The results for untrained and trained participants show no significant difference for non-pooled stimuli (Table 3) as well as the pools for tonality, t_{tonal} (35.91) = -0.27, p = .788, and complexity, t_{atonal} (32.75) = 0.55, p = .586. All differences were again tested with pairwise t-tests with Benjamini-Hochberg-corrections (Benjamini & Hochberg, 2000) if necessary. Similarly, TJND differences between musically active and inactive participants were also not significant, t_{active} (16.46) = -0.11, p = .917, and the same is the case for gender differences, t_{gender} (13.23) = 0.16, p = .874.

5.2.2 Table 3

Group comparisons between untrained and trained participants.

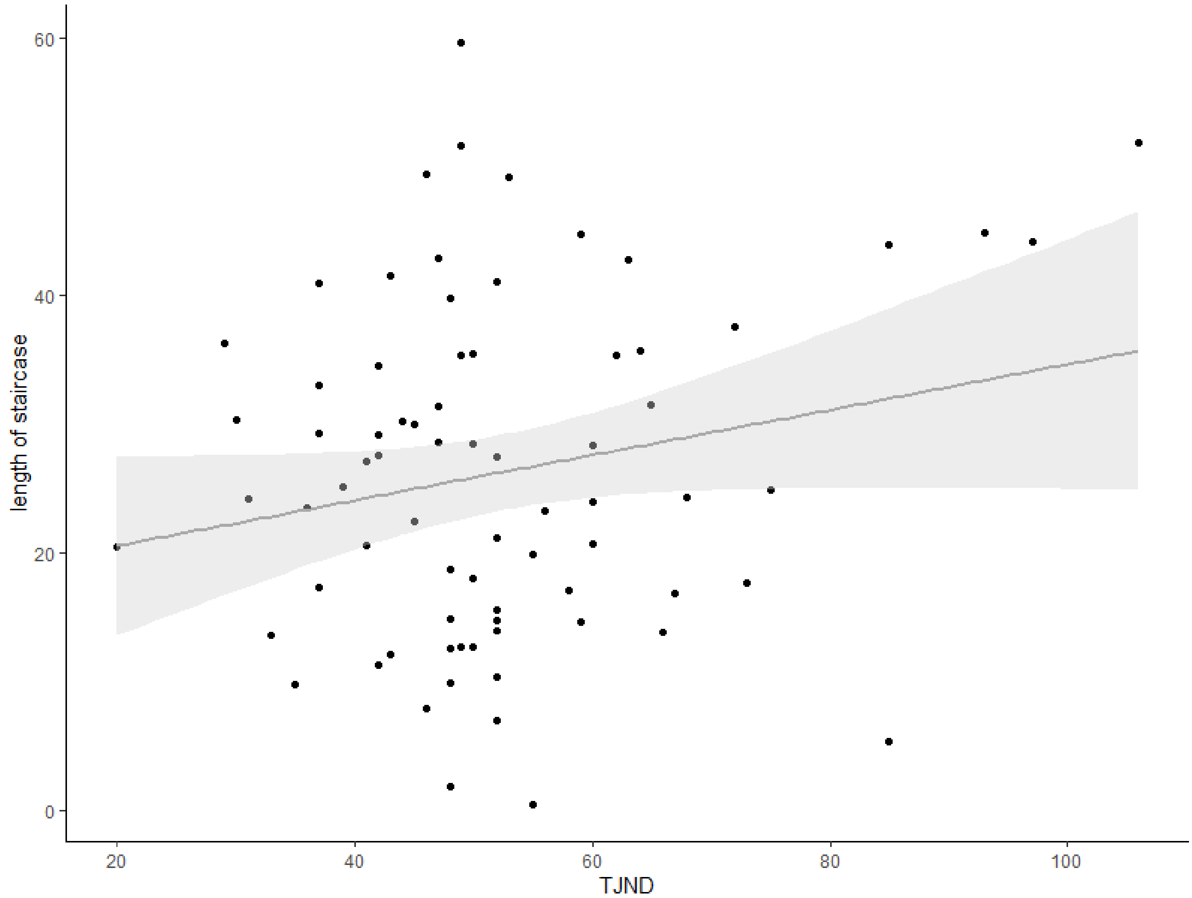
	Overall $\bar{x}$ (ms)	SD	Untrained $\bar{x}$ (ms)	SD	Trained $\bar{x}$ (ms)	SD	DF	t	p
LC	11.66	5.70	10.77	6.53	12.65	4.80	16.39	-0.72	.717
HD	41.65	8.38	41.13	9.31	42.23	7.7	16.90	-0.28	.782
LA	27.11	9.83	29.37	7.75	24.61	11.69	13.67	1.03	.717
HA	24.53	6.04	23.70	6.04	25.48	6.25	16.65	-0.62	.717

Note. Comparison of TJND-means of participant groups of all stimuli groups. *p*-values are corrected by Benjamini-Hochberg-Method.

Analysis of the staircase-length showed a noticeable but insignificant trend towards a higher TJND with increasing trials in the staircase, $r\,(74) = 1.81$, $p = .074$ (Figure 4). There was also no difference in the number of trials per stimuli. A post-hoc power analysis with the WebPower-package in R for the repeated measures ANOVA resulted in a power of 81 % for within-effect testing, showing that the initial power analysis by bootstrapping the first four results was an accurate estimator of the whole sample. Initially, there was some doubt about the data quality due to the large differences in variance and variability of the results, which were not expected to be this high. But as the results of the ANOVA could be replicated with different statistical procedures which are more robust to differences in variance, the variability in the results cannot be attributed to the data quality but rather to group differences.

5.2.3 Figure 4

Scatterplot with the staircase length and final TJND.



Note. Most points accumulate around the average staircase length and TJND with no clear tendency in either direction. The average trend going slightly upwards can be attributed to four outliers, where participants took a particularly long time to finish the staircase due to the experimental setup. Inspection of these staircases showed that turnarounds were often not counted, as they only were counted when participants had the same result two times in a row, resulting in a prolonged procedure with no significant change in TJND. Possible changes to the procedure are discussed further down.

Overall, our results do not show tonal equivalence, as the results of the non-pooled stimuli are in line with this expectation, but this effect disappears in the pooled results and contradicts the other results. In detail, we found no significant difference between tonal and atonal stimuli, thus rejecting the tonality condition (H1a), while the complexity condition is accepted (H1b), as the difference between low complex and high complex stimuli is significantly different with a moderate effect size of d_{comp} = .65, meaning that participants have lower TJND with low-complex stimuli. With non-pooled results, we accept H1c as the difference between LC and HD is significant with an exceptionally strong effect size of $d_{LC\text{-}HD}$ = 2.52. But it also needs to be noted that the standard deviation of HD (16.76) is double the amount of LC (8.15), which could influence the conducted test, as Mauchly's Test of Sphericity was only just nonsignificant and thus the statistical results could be exaggerated. On the other hand, the difference between the control stimuli HA and HD is insignificant, rejecting H1d in line with our expectations.

For this sample, we observed no significant effect for all demographic variables and measures of musical training, as all inter-stimuli differences for untrained and trained participants were insignificant. But we need to emphasize that further implementation of this paradigm with more participants could lead to significant differences.

6. Discussion

In the experiment we tested the performance of participants to differentiate between two nearly identical polyrhythms that were either tonal or atonal and had low or high complexity. We found that only the TJND of the tonal stimuli were significantly different, while mean and variance of the atonal stimuli were nearly identical. The differences between LC and HD contribute to the significant result in the complexity condition, where LC skews the low complexity grouping down and HD skews the high complexity grouping upwards (Figure 3C). When pooling both stimuli in the tonality condition (Figure 3B) the observed effect completely vanishes with a nonsignificant result against the atonal grouping, which conflicts with the group results, especially the high TJND of HD. Before we discuss possible reasons for these results further, we can confirm that the measures of avoiding training effect by randomizing tempo and interleaving the trial order were effective as we see no trend downwards over time (Figure 4).

6.1 The Difference of LC and HD

The experiment failed to show conclusive evidence for the proposed similarity of rhythm and pitch processing concerning C/D. While the stark difference in performances of LC and HD can be seen as an effect of C/D, there are some other possible factors likely at play that contribute to the effect. The most obvious explanation for LC performing best is salience due to rhythm-familiarity and the lack of relative complexity. Firstly, rhythm-familiarity is based on the exposure of this polyrhythm in popular music, where it is frequently used, and would lead to a processing advantage, since the internal model does not have to learn a completely unfamiliar rhythm (Stupacher et al., 2017). Secondly, we assumed that the intimate familiarity with tone intervals would translate to LC and HD, which is so strong in pitch perception that slightly deviant intervals are categorized as neighboring tone intervals, suggesting a strong internal memorization of the 12-TET system (Bailes et al., 2015). As this is also the case for dissonant intervals (as HD would be), this should enhance the predictive precision and thus lead to better performance to atonal stimuli, since participants were not completely unfamiliar with them. But since HD has performed worse than even the atonal stimuli, we can deduce that tone-familiarity had no effect on HD and indicates that our initial assumption of similarity can be at least partly discarded.

The lack of relative complexity of LC compared to the others, which were all theoretically very complex and thus demanding, is likely another contributing factor towards its better performance. The results of Møller et al. (2021) show a stark contrast between the polyrhythms 2:3 (LC) and 3:5 (LA) in the amount of indistinguishable patterns produced by finger tapping, suggesting a stark advantage in predictive power of the internal model for LC. The study was published after the design of the experiment was finalized so we could not incorporate these findings into the experimental design. When we extrapolate the results towards the high-complex stimuli in our experiment, we can conclude that the cognitive load for the internal model for all stimuli excluding LC was very high. We can assume that the inverse-U-relationship of complexity and pleasure for complex rhythms proposed by Vuust and Witek (2014), where pleasure is a representation of the learning effect of the internal model from prediction errors, does not apply in our case, as the complexity of our stimuli is off the scale. Unfortunately, we are unable to calculate any reference of how complex each stimuli truly was, as the use of mathematical estimations like Shannon entropy, which is a measure of information content for possible outcomes of a variable (Basiński et al., 2023; Lumaca et al., 2019; Shannon, 1948), were not considered, as we instead chose the numerical complexity of ratios as the estimator of complexity. Although familiarity and complexity influence the performance of LC, we cannot estimate the size of influence on the results and thus can neither fully dismiss nor fully accept that LC's performance can be attributed to its tonal equivalence as HD is also equivalent to a tone interval.

In this respect, it is especially interesting that HD performed worst overall. Like LC, familiarity should have enhanced HD's performance, although to a lesser degree then LC, if the tonality of HD was of any influence. As we have established based on the results of Møller et al. (2021), there should not be a sizeable difference between HD and HA considering that both are seemingly too complex for an untrained internal model to make predictions apart from guessing (Vuust & Witek, 2014). If we assume tone-equivalence then HD could perform worse, because dissonant intervals take slightly longer to process due to a higher cognitive load (Andermann et al., 2020) which leads to worse precision as the internal model already struggles to find any pattern. This could also explain the difference between HD and HA, since the internal model is not familiar with HA and thus the cognitive load would be high but reduced compared to HD. In this case, the tone-familiarity of HD would be detrimental to the performance, if this is indeed the case, which we cannot prove with this experiment since HD is the only dissonant stimuli.

The similar results of LA and HD suggest a floor effect for processing precision for very complex rhythms. Although it is likely that the internal model was not able to establish a meter for these polyrhythms, the comparative nature of the experiment enabled participants to keep the regularity of it in short memory for a direct comparison as the pause between stimuli was only 400ms. If this was the case, then the TJND should be higher when the pause is longer, as the memory should get less precise over time (Stupacher et al., 2017). We now take a deeper look at grouping effects.

6.2 Grouping Effects

The intention for pooling the stimuli by tonality and complexity was to determine if an effect can be attributed to one of these correlated factors and thus separating both highly correlated factors. While the achieved this goal since we observed an effect for complexity but not for tonality, it is not as clear cut as it seems at first glance. The complexity-effect is moderate to large (d_{comp} = .65), but the main contributors to the effect are only differences in the tonal stimuli. The atonal stimuli had quite similar results in contrast, so there is merit to the argument that the complexity-effect simply has its roots in tonality-effects discussed before, even though an effect in the tonality-pool was not found. Following this line of argument, possible explanations like familiarity and complexity would also apply here, since LC could have a precision advantage due to it being more familiar and less complex.

With our limited knowledge, this interpretation makes the most sense as the hypothesis of differences due to complexity is flawed as newer evidence shows that our categorization based on numerical complexity is too reductive. The study by Møller et al. (2021) used a finger-tapping-task with equivalent polyrhythms and only 41 out of 120 participants in their ratio-experiment showed a recognizable finger-tapping pattern in the slowest conditions for LA, while over 80 participants could tap a recognizable pattern for LC in the same condition. For faster conditions, the results are even worse for LA. This shows that LA cannot be labeled as low complex and as such, the comparison between low and high complexity groupings is nonsensical.

The pooling of tonal and atonal stimuli is valid due to it being based on established musical concepts. Here we observed no effect, which contradicts the results found on group level. It is statistically remarkable that the means of the tonal and atonal pool are so similar ($\bar{x}_{tonal}$= 26.6ms, $\bar{x}_{atonal}$= 25.8), but as the variances are reflecting the different distributions of LC and HD (SD_{tonal}= 16.76, SD_{atonal}= 8.15), it is likely that that this has no further meaning.

In sum, the results of this experiment show a significant result between LC and HD, which cannot be fully explained due to shortcomings in the design. Likely explanations are the hypothesized effect of processing differences between C/D, familiarity and complexity, but no explanation can be fully dismissed and thus the discussion remains inconclusive. We now discuss these shortcomings of the experiment and propose improvements to the design.

6.3 Limitations and Future Research

While the experiment was successful in showing a strong effect in the predicted direction, we cannot conclusively attribute this effect to similarities in processing of rhythm and pitch. There are indications that the processing advantage of consonant intervals may also apply to rhythm processing, as suggested by the performance of LC, which could be attributed to a higher precision in the internal model. But due to limitations in the data and the fact that this experiment is the first of its kind, we cannot estimate the influence of other factors like familiarity with the current setup, where the main goal was to entangle the relationship between C/D and complexity.

From our understanding at the design stage, there was no indication that HD should be the worst performing stimuli and as such, we did not account for this occurrence, which was a critical oversight in the design. With the amount of data present, we are unable to entangle the contradictions in our results. To remedy this, future experiments should include more tonal polyrhythms. Similarly, the assessment of complexity for polyrhythms was reductive and, as later studies revealed (Møller et al., 2021), not a valid categorization system and the homogenous sample with regards to musical training made controlling for it difficult as not all participant groups were sufficiently large.

Regarding the experiment itself, the use of turnarounds for estimation of the TJND was sufficient as later turnarounds had consistent JF indicating a reliable measure, but they are inefficient, as it took participants often a considerable number of trials to finish a staircase even though there was little change in JF. In most cases, this happened when they were not fulfilling the criteria for a turnaround (two consecutive correct/wrong answers in a row) and were thus stuck in a loop with no significant information gained. Lastly, the number of stimuli was deliberately chosen due to time and concentration constraints (due to staircase length) and to minimize the complexity of the experiment to only essential elements without unnecessary complications. But with hindsight, this reduction hampered the exclusion of alternative explanations, as we cannot say if the results of HD can also be found in other dissonant polyrhythms, because we do not have the data. We propose that future experiments with this paradigm should incorporate the following changes:

1. A more selective process for sampling to ensure an equal distribution for musical training or focus on musicians. The assessment of training should not only be self-reported but based on more objective tests like Ollen Musical Sophistication Index questionnaire (Ollen, 2006). This should further enable the control of musical training in the experiment.

2. Estimating the TJND using Bayesian adaptive testing (Kuravsky et al., 2017) during trials instead of counting the number of turnarounds. This should reduce staircase length and thus cognitive load over time on participants, which should also enable the use of more stimuli in one session.
3. Using the whole tone scale for stimuli generation. This enables a better comparison of C/D and other literature such as Itoh et al. (2010), who used all tone intervals in an octave, which enabled a better comparison of C/D. But in this approach, stimuli length must be considered, as a diminished fifths with a ratio of 1024:729 would not complete a whole bar in the time we used in our design, which could confound the analysis.
4. Using a larger sample size. Although our sample was sufficient for the effect we found, we highly suggest a larger sample size to further the differences we found.

7. Conclusion

This experiment demonstrated a clear distinction between consonant and dissonant polyrhythms in this experimental paradigm. However, due to the nature of the results, we are unable to confirm whether tonality was the primary factor contributing to the observed effect, or if other variables may have played a role, particularly in the case of the consonant polyrhythm. The poorer performance of the dissonant polyrhythm compared to atonal polyrhythms raises more questions than it answers, as no current model of rhythm or pitch processing can fully explain this outcome. We strongly encourage future research to investigate this result by comparing it with other dissonant polyrhythms to determine if it can be reproduced. If confirmed, these findings could challenge the current understanding of rhythm processing. If this result is confirmed, it could challenge and expand the current understanding of rhythm processing. This exploratory experiment ventured into the fringes of rhythm processing models by investigating rhythms previously considered too complex to study without the context of C/D (Møller et al., 2021). While this experiment uncovered intriguing findings, their full implications to PCM and the research of pitch and rhythm remain uncertain.

8. References

Andermann, M., Patterson, R. D., & Rupp, A. (2020). Transient and sustained processing of musical consonance in auditory cortex and the effect of musicality. *Journal of Neurophysiology*, *123*(4), 1320–1331. https://doi.org/10.1152/jn.00876.2018

Attneave, F., & Olson, R. K. (1971). Pitch as a Medium: A New Approach to Psychophysical Scaling. *The American Journal of Psychology*, *84*(2), 147. https://doi.org/10.2307/1421351

Bailes, F., Dean, R. T., & Broughton, M. C. (2015). How Different Are Our Perceptions of Equal-Tempered and Microtonal Intervals? A Behavioural and EEG Survey. *PloS One*, *10*(8), e0135082. https://doi.org/10.1371/journal.pone.0135082

Bain, R. (2003). *The harmonic series*. University of South Carolina School of Music. https://anneloeswolters.com/wp-content/uploads/2019/08/overtone_series_1.pdf

Basiński, K., Quiroga-Martinez, D. R., & Vuust, P. (2023). Temporal hierarchies in the predictive processing of melody - From pure tones to songs. *Neuroscience and Biobehavioral Reviews*, *145*, 105007. https://doi.org/10.1016/j.neubiorev.2022.105007

Benjamini, Y., & Hochberg, Y. (2000). On the Adaptive Control of the False Discovery Rate in Multiple Testing With Independent Statistics. *Journal of Educational and Behavioral Statistics*, *25*(1), 60–83. https://doi.org/10.3102/10769986025001060

Bidelman, G. M. (2013). The role of the auditory brainstem in processing musically relevant pitch. *Frontiers in Psychology*, *4*, 264. https://doi.org/10.3389/fpsyg.2013.00264

Bidelman, G. M., & Krishnan, A. (2009). Neural correlates of consonance, dissonance, and the hierarchy of musical pitch in the human brainstem. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *29*(42), 13165–13171. https://doi.org/10.1523/JNEUROSCI.3900-09.2009

Brugge, J. F., Nourski, K. V., Oya, H., Reale, R. A., Kawasaki, H., Steinschneider, M., & Howard, M. A. (2009). Coding of repetitive transients by auditory cortex on Heschl's gyrus. *Journal of Neurophysiology*, *102*(4), 2358–2374. https://doi.org/10.1152/jn.91346.2008

Burred, J. J., Robel, A., & Sikora, T. (2010). Dynamic Spectral Envelope Modeling for Timbre Analysis of Musical Instrument Sounds. *IEEE Transactions on Audio, Speech, and Language Processing*, *18*(3), 663–674. https://doi.org/10.1109/tasl.2009.2036300

Carcagno, S., Lakhani, S., & Plack, C. J. (2019). Consonance perception beyond the traditional existence region of pitch. *The Journal of the Acoustical Society of America*, *146*(4), 2279. https://doi.org/10.1121/1.5127845

Chan, P. Y., Dong, M., & Li, H. (2019). The Science of Harmony: A Psychophysical Basis for Perceptual Tensions and Resolutions in Music. *Research (Washington, D.C.)*, *2019*, 2369041. https://doi.org/10.34133/2019/2369041

Cheveigné, A. de. (2005). Pitch Perception Models. In C. J. Plack, A. J. Oxenham, R. R. Fay, & A. N. Popper (Eds.), *Springer Handbook of Auditory Research: Vol. 24. Pitch: Neural coding and perception* (pp. 169–233). Springer. https://doi.org/10.1007/0-387-28958-5_6

Ciria, A., Schillaci, G., Pezzulo, G., Hafner, V. V., & Lara, B. (2021). Predictive Processing in Cognitive Robotics: A Review. *Neural Computation*, *33*(5), 1402–1432. https://doi.org/10.1162/neco_a_01383

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology*, *1*(1), 42–45. https://doi.org/10.20982/tqmp.01.1.p042

Crespo-Bojorque, P., Monte-Ordoño, J., & Toro, J. M. (2018). Early neural responses underlie advantages for consonance over dissonance. *Neuropsychologia*, *117*, 188–198. https://doi.org/10.1016/j.neuropsychologia.2018.06.005

Di Stefano, N., Vuust, P., & Brattico, E. (2022). Consonance and dissonance perception. A critical review of the historical sources, multidisciplinary findings, and main hypotheses. *Physics of Life Reviews*, *43*, 273–304. https://doi.org/10.1016/j.plrev.2022.10.004

Ebeling, M. (2008). Neuronal periodicity detection as a basis for the perception of consonance: A mathematical model of tonal fusion. *The Journal of the Acoustical Society of America*, *124*(4), 2320–2329. https://doi.org/10.1121/1.2968688

Friston, K. (2008). Hierarchical models in the brain. *PLoS Computational Biology*, *4*(11), e1000211. https://doi.org/10.1371/journal.pcbi.1000211

Hallam, S., Cross, I., Thaut, M., & Jones, M. R. (2012). *Musical time* (Vol. 1). Oxford University Press. https://academic.oup.com/edited-volume/35473/chapter/303809183 https://doi.org/10.1093/oxfordhb/9780199298457.013.0008

Heggli, O. A., Konvalinka, I., Kringelbach, M. L., & Vuust, P. (2019). Musical interaction is influenced by underlying predictive models and musical expertise. *Scientific Reports*, *9*(1), 11048. https://doi.org/10.1038/s41598-019-47471-3

Helmholtz, H. von. (1865). *Die Lehre Von Den Tonempfindungen Als Physiologische Grundlage Für Die Theorie der Musik* (6th ed.). Springer Vieweg. in Springer Fachmedien Wiesbaden GmbH.

Hyde, K. L., Lerch, J., Norton, A., Forgeard, M., Winner, E., Evans, A. C., & Schlaug, G. (2009). Musical training shapes structural brain development. *Journal of Neuroscience*, *29*(10), 3019–3025. https://doi.org/10.1523/JNEUROSCI.5118-08.2009

Itoh, K., Suwazono, S., & Nakada, T. (2010). Central auditory processing of noncontextual consonance in music: An evoked potential study. *The Journal of the Acoustical Society of America*, *128*(6), 3781–3787. https://doi.org/10.1121/1.3500685

Kumar, S., & Schönwiesner, M. (2012). Mapping human pitch representation in a distributed system using depth-electrode recordings and modeling. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *32*(39), 13348–13351. https://doi.org/10.1523/JNEUROSCI.3812-12.2012

Kumar, S., Sedley, W., Nourski, K. V., Kawasaki, H., Oya, H., Patterson, R. D., Howard, M. A., Friston, K. J., & Griffiths, T. D. (2011). Predictive coding and pitch processing in the auditory cortex. *Journal of Cognitive Neuroscience*, *23*(10), 3084–3094. https://doi.org/10.1162/jocn_a_00021

Kuravsky, L. S., Artemenkov, S. L., Yuryev, G. A., & Grigorenko, E. L. (2017). A new approach to computerized adaptive testing. *Experimental Psychology (Russia)*, *10*(3), Article 3, 33–45. https://doi.org/10.17759/exppsy.2017100303

Lahdelma, I., & Eerola, T. (2020). Cultural familiarity and musical expertise impact the pleasantness of consonance/dissonance but not its perceived tension. *Scientific Reports*, *10*(1), 8693. https://doi.org/10.1038/s41598-020-65615-8

Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review*, *106*(1), 119–159. https://doi.org/10.1037//0033-295x.106.1.119

Li, C.-W., Guo, F.-Y., & Tsai, C.-G. (2021). Predictive processing, cognitive control, and tonality stability of music: An fMRI study of chromatic harmony. *Brain and Cognition*, *151*, 105751. https://doi.org/10.1016/j.bandc.2021.105751

Livshin, A., & Rodet, X. (2006). The Significance of the Non-Harmonic “Noise” Versus the Harmonic Series for Musical Instrument Recognition. In *International Symposium on Music Information Retrieval (ISMIR’06)*. https://hal.science/hal-01161372/

London, J. (2012). *Hearing in Time: Psychological Aspects of Musical Meter*. Oxford University Press.

Lumaca, M., Trusbak Haumann, N., Brattico, E., Grube, M., & Vuust, P. (2019). Weighting of neural prediction error by rhythmic complexity: A predictive coding account using mismatch negativity. *The European Journal of Neuroscience*, *49*(12), 1597–1609. https://doi.org/10.1111/ejn.14329

Majdak, P., & Mihocic, M. (2020). *ExpSuite* (Version 1.1.52) [Computer software]. Acoustics Research Institute of the Austrian Academy of Sciences. https://www.oeaw.ac.at/isf/expsuite

McDermott, J. H., Lehr, A. J., & Oxenham, A. J. (2010). Individual differences reveal the basis of consonance. *Current Biology : CB*, *20*(11), 1035–1041. https://doi.org/10.1016/j.cub.2010.04.019

Millidge, B., Seth, A., & Buckley, C. L. (2021, July 27). *Predictive Coding: a Theoretical and Experimental Review*. http://arxiv.org/pdf/2107.12979v4

Møller, C., Stupacher, J., Celma-Miralles, A., & Vuust, P. (2021). Beat perception in polyrhythms: Time is structured in binary units. *PloS One*, *16*(8), e0252174. https://doi.org/10.1371/journal.pone.0252174

Neely, A. (2018, March 5). *Polyrhythms are Polypitch - live @ Ableton Loop, Berlin* [Video]. Ableton Loop. Youtube, Berlin. https://www.youtube.com/watch?v=-tRAkWaeepg

Nourski, K. V., Brugge, J. F., Reale, R. A., Kovach, C. K., Oya, H., Kawasaki, H., Jenison, R. L., & Howard, M. A. (2013). Coding of repetitive transients by auditory cortex on posterolateral superior temporal gyrus in humans: An intracranial electrophysiology study. *Journal of Neurophysiology*, *109*(5), 1283–1295. https://doi.org/10.1152/jn.00718.2012

Nozaradan, S., Schönwiesner, M., Keller, P. E., Lenc, T., & Lehmann, A. (2018). Neural bases of rhythmic entrainment in humans: Critical transformation between cortical and lower-level representations of auditory rhythm. *The European Journal of Neuroscience*, *47*(4), 321–332. https://doi.org/10.1111/ejn.13826

Oberfeld, D., & Franke, T. (2013). Evaluating the robustness of repeated measures analyses: The case of small sample sizes and nonnormal data. *Behavior Research Methods*, *45*(3), 792–812. https://doi.org/10.3758/s13428-012-0281-2

Ollen, J. E. (2006). *A criterion-related validity test of selected indicators of musical sophistication using expert ratings: PhD thesis* [, The Ohio State University]. rave.ohiolink.edu. https://rave.ohiolink.edu/etdc/view?acc_num=osu1161705351

Oxenham, A. J. (2023). Questions and controversies surrounding the perception and neural coding of pitch. *Frontiers in Neuroscience*, *16*, 1074752. https://doi.org/10.3389/fnins.2022.1074752

Oxenham, A. J., Micheyl, C., Keebler, M. V., Loper, A., & Santurette, S. (2011). Pitch perception beyond the traditional existence region of pitch. *Proceedings of the National Academy of Sciences*, *108*(18), 7629–7634. https://doi.org/10.1073/pnas.1015291108

Pallesen, K. J., Bailey, C. J., Brattico, E., Gjedde, A., Palva, J. M., & Palva, S. (2015). Experience Drives Synchronization: The phase and Amplitude Dynamics of Neural Oscillations to Musical Chords Are Differentially Modulated by Musical Expertise. *PloS One*, *10*(8), e0134211. https://doi.org/10.1371/journal.pone.0134211

Park, J. Y., Park, H., Kim, J.-I., & Park, H.-J. (2011). Consonant chords stimulate higher EEG gamma activity than dissonant chords. *Neuroscience Letters*, *488*(1), 101–105. https://doi.org/10.1016/j.neulet.2010.11.011

Pearce, M. T [M. T.]. (2005). *The construction and evaluation of statistical models of melodic structure in music perception and composition*. openaccess.city.ac.uk. https://openaccess.city.ac.uk/id/eprint/8459/

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87. https://doi.org/10.1038/4580

Rohrmeier, M. A., & Koelsch, S. (2012). Predictive information processing in music cognition. A critical review. *International Journal of Psychophysiology : Official Journal of the International Organization of Psychophysiology*, *83*(2), 164–175. https://doi.org/10.1016/j.ijpsycho.2011.12.010

Rufener, K. S., Husemann, A.-M., & Zaehle, T. (2020). The internal time keeper: Causal evidence for the role of the cerebellum in anticipating regular acoustic events. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *133*, 177–187. https://doi.org/10.1016/j.cortex.2020.09.021

Salimpoor, V. N., Zald, D. H., Zatorre, R. J., Dagher, A., & McIntosh, A. R. (2015). Predictions and the brain: How musical sounds become rewarding. *Trends in Cognitive Sciences*, *19*(2), 86–91. https://doi.org/10.1016/j.tics.2014.12.001

Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, *27*(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Strait, D. L., Kraus, N., Parbery-Clark, A., & Ashley, R. (2010). Musical experience shapes top-down auditory mechanisms: Evidence from masking and auditory attention performance. *Hearing Research*, *261*(1-2), 22–29. https://doi.org/10.1016/j.heares.2009.12.021

Stupacher, J. (2019). The experience of flow during sensorimotor synchronization to musical rhythms. *Musicae Scientiae*, *23*(3), 348–361. https://doi.org/10.1177/1029864919836720

Stupacher, J., Wood, G., & Witte, M. (2017). Neural Entrainment to Polyrhythms: A Comparison of Musicians and Non-musicians. *Frontiers in Neuroscience*, *11*, 208. https://doi.org/10.3389/fnins.2017.00208

Tabas, A., Andermann, M., Schuberth, V., Riedel, H., Balaguer-Ballester, E., & Rupp, A. (2019). Modeling and MEG evidence of early consonance processing in auditory cortex. *PLoS Computational Biology*, *15*(2), e1006820. https://doi.org/10.1371/journal.pcbi.1006820

Tenney, J. (1988). *A history of "consonance" and "dissonance"*. Excelsior Music Publ. Co.

Tramo, M. J., Cariani, P. A., Delgutte, B., & Braida, L. D. (2001). Neurobiological Foundations for the Theory of Harmony in Western Tonal Music. *Annals of the New York Academy of Sciences*, *930*(1), 92–116. https://doi.org/10.1111/j.1749-6632.2001.tb05727.x

Ungan, P., & Yagcioglu, S. (2014). Significant variations in Weber fraction for changes in inter-onset interval of a click train over the range of intervals between 5 and 300 ms. *Frontiers in Psychology*, *5*, 1453. https://doi.org/10.3389/fpsyg.2014.01453 van der Weij, B., Pearce, M. T [Marcus T.], & Honing, H. (2017). A Probabilistic Model of Meter Perception: Simulating Enculturation. *Frontiers in Psychology*, *8*, 824. https://doi.org/10.3389/fpsyg.2017.00824

Verschooten, E., Shamma, S., Oxenham, A. J., Moore, B. C. J., Joris, P. X., Heinz, M. G., & Plack, C. J. (2019). The upper frequency limit for the use of phase locking to code temporal fine structure in humans: A compilation of viewpoints. *Hearing Research*, *377*, 109–121. https://doi.org/10.1016/j.heares.2019.03.011

Virtala, P., & Tervaniemi, M. (2017). Neurocognition of Major-Minor and Consonance-Dissonance. *Music Perception*, *34*(4), 387–404. https://doi.org/10.1525/mp.2017.34.4.387

Vuust, P., Dietz, M. J., Witek, M., & Kringelbach, M. L. (2018). Now you hear it: A predictive coding model for understanding rhythmic incongruity. *Annals of the New York Academy of Sciences*. Advance online publication. https://doi.org/10.1111/nyas.13622

Vuust, P., Heggli, O. A., Friston, K. J., & Kringelbach, M. L. (2022). Music in the brain. *Nature Reviews. Neuroscience*, *23*(5), 287–305. https://doi.org/10.1038/s41583-022-00578-5

Vuust, P., & Witek, M. A. G. (2014). Rhythmic complexity and predictive coding: A novel approach to modeling rhythm and meter perception in music. *Frontiers in Psychology*, *5*, 1111. https://doi.org/10.3389/fpsyg.2014.01111

Walsh, K. S., McGovern, D. P., Clark, A., & O'Connell, R. G. (2020). Evaluating the neurophysiological evidence for predictive processing as a model of perception.

Annals of the New York Academy of Sciences, *1464*(1), 242–268. https://doi.org/10.1111/nyas.14321

Yuan, K.-H., & Hayashi, K. (2003). Bootstrap approach to inference and power analysis based on three test statistics for covariance structure models. *The British Journal of Mathematical and Statistical Psychology*, *56*(Pt 1), 93–110. https://doi.org/10.1348/000711003321645368

Zhang, Z., & Yuan, K.-H. (2018). *Practical statistical power analysis using Webpower and R*. ISDSA Press. https://doi.org/10.35566/power

Zhao, T. C., Lam, H. T. G., Sohi, H., & Kuhl, P. K. (2017). Neural processing of musical meter in musicians and non-musicians. *Neuropsychologia*, *106*, 289–297. https://doi.org/10.1016/j.neuropsychologia.2017.10.007

9. Appendix: Questionnaire for musical training

1. Are you musically active or were you in the past?
2. Do you play an instrument? If yes, for how many years have you been playing it?
3. Did you ever play a rhythmical instrument over a prolonged period of time e.g., drums or cajon?
4. On a scale of 1 to 7, how well would you evaluate your ability to follow a given rhythm e.g., through clapping?
5. For how many years did you have formal musical training in the past?
6. Do you have a musical degree? If yes, in which domain?