Spring 2021

The Role of a Polyrhythm's Pitch Interval in Music-Dependent Memory

Hadley R. Parum
Bard College

The Role of a Polyrhythm's Pitch Interval in Music-Dependent Memory

Senior Project Submitted to

The Division of Science, Math, and Computing & The Division of The Arts

of Bard College

by

Hadley Parum

Annandale-on-Hudson, New York

May, 2021

Acknowledgements

So many friends, mentors, and colleagues have had profound impacts on my approach to work and life; in the pages of this project there are certainly pieces of people whose names will not be below, and this shouldn't discredit the love I have for them, though it remains out of print. The challenge of this section will be to thank as many people who have directly contributed to the work on this project over (more than) this past year, and I promise to voice thanks and/or disperse hugs at the request of anyone else.

Dr. Jacko, it's your fault I'm at Bard, and it's thanks to you that music has remained a part of my life at all. I am always sending you and Jayci my love, and I look forward to whenever next we can all brunch together.

Mrs. Reitz, I have you to thank for my entry into Psychology, and for modelling a no-bullshit approach to learning and life that has been a catalyst for embracing my own passion everywhere I find it. Many of your words have become my mantras, and damn good ones.

Justin Hulbert, my surviving this project and these past four years is in no small part thanks to you. You have been a fantastic advisor at all stages of these past years: you have eased panic, laughed through all of the gripes, encouraged my best habits, and provided more support than any one human should be allowed to provide.

Kyle Gann, I have been grateful for all of our conversations, and I am glad to have had you as my advisor during this project's journey. Thank you for putting up with all directions of my enthusiasm, and I hope that I've cured at least one instance of Columbia White Boy Syndrome. If you ever need someone to come laugh at your jokes, you know where to find me.

Sarah Dunphy-Lelii, without your email at the end of 2019, this project would not have existed in this form at all, and I'm unsure my last year at Bard would have had the space for joy in all the places I was able to find it. For your confidence and kindness, thank you.

John Cullinan, I am forever indebted to you for your caffeinic, emotional, and intellectual support. You remind me of where I find love within and outside of all the projects I've packed into this thing called life, and I hope I've reciprocated at least some fraction of the good tunes, talks, and tasty treats. No one reads them, but I dedicate the liner notes to you.

Dad, I blame you for most of who I am. But, the joke's on you: you're stuck with me! Let's hope I'm not too rusty as your backup next time we play. I love talking to you, and I'm thankful for your stories, jokes (even at my expense), and suggestions (even when I don't take them). I look forward to talking your ear off about this project at your earliest convenience.

Additionally, this project could not have been completed without financial support from the Psychology program, John Cullinan, the Lifetime Learning Institute, and Psi Chi, the Honor Society for Psychology.

Digital Materials

This Senior Project would be incomplete without the music that has informed it and is referenced within. For the reader's listening pleasure, I have compiled a YouTube playlist with the pieces of music referenced in the project, which they can listen to by either clicking these words or visiting https://tinyurl.com/parumsounds in their browser of choice.

Additionally, practicing transparency when conducting scientific research has been made easier by the Open Science Foundation. While some materials are included in the Appendices of this project, a more complete and up-to-date record of preregistrations, materials, and code are available at the project's repository here, at https://osf.io/235xb/.

Table of Contents

Abstract

When listening to music, humans can easily and often automatically assess the perceptual similarity of different moments in music. However, it is difficult to rigorously define the way in which we determine exactly how similar we find moments to be. This problem has driven inquiry in music cognition, musicology, and music theory alike, but previous results have depended on behaviorally mediated responses and/or recursive analytic strategies by music scholars. The present work employs the context-dependent memory paradigm as a novel way to investigate the extent to which listeners consider two musical examples to be similar. After incidentally learning words while listening to a 5:4 polyrhythm forming a perfect fifth, participants could hear no sound or the polyrhythm at a different pitch interval during a surprise test of recall. Between-subjects comparisons found no effect of the actual sound context at test on recall; however, participants who reported being in the same sound context did recall significantly more words than others. Interactions between actual and reported sound context were not accounted for by musical experience or other participant factors, and reported sound context was more often incompatible than compatible with actual sound context. Contributions to mental context theory and the boundaries of conclusions about musical features are discussed.

Keywords: context-dependent memory, free recall, perception, pitch, memory, music cognition, rhythm

It is a trivial challenge for most humans to tell whether they are listening to a Beethoven Symphony or a Balinese Gamelan ensemble, and for good reason. At the same time, it is also relatively common for a moment in a piece of music to remind the listener of something they've heard before. The bassline of a new pop song may sound like a common lullaby, or the penultimate song on an album may bring back the melody from the opening track. This can enrich the experience of listening to music, as connections between new experiences and music become more intricate webs, and new music finds ways to cue old memories. In fact, musicians often rely on these connections: jazz musicians quote well-known melodies that can be familiar to audiences even when transposed and heavily embellished; film scores are rife with themes that exemplify characters, or cue viewers into an imminent fight scene; long orchestral works spend hours exploring different forms of some musical idea, and deliver satisfaction to listeners by returning to a familiar melody that was established in the piece's first minute.

Musicians who, intentionally or unintentionally, employ musical ideas in their work deemed too similar to another artists' work can find themselves facing practical consequences. As one prominent example, the 2013 song "Dark Horse," performed by pop artist Katy Perry, earned her a lawsuit in 2014 in which the Christian rapper Flame accused her of plagiarizing a melodic ostinato from his song "Joyful Noise," released in 2008. In 2019, a jury ruled in favor of Flame, but Perry won an appeal the following year (Blistein, 2020). An online resource sponsored by the law schools of George Washington University and Columbia University catalogues the increasing number of music copyright claims that have made it to court in the past decade (*Cases from 2010-2019,* n.d.). While music copyright claims more often plague digital content creators playing portions of published songs, these inter-artist disputes about the origin of certain basslines, melodies, or chord patterns have become notorious and commonly reported (Mullan, n.d.; Wang & Wang, 2020).

There has also been incredible sociopolitical weight to music sounding like certain other music, or ideas of musical moments. The incorporation of folk song into classical musical practices is one way that musicians have attempted to write music that sounds authentically of their nation, even as this freezing of pastoral or pre-colonial aesthetics has also created standards of authenticity that do not evolve as people do. While a review of this phenomenon is a Senior Project in its own right, I encourage readers to explore the transformation of folk music into classical traditions in the formation of the Soviet Union (Frolova-Walker, 1998; Levin, 2013; Rothstein, 1980) and early Communist China (Jones, 2001; Mu, 1994; Tse-tsung, 1956), as well as by composers like Béla Bartók of Hungary (Suchoff, 1972; Tari, 2006), Ralph Vaughan Williams of England (Kimmel, 1941; Williams et al., 1906; I highly recommend his *Fantasia on a Theme by Thomas Tallis*), as well as much of American popular music.

Musicians frequently quote or employ motifs from recognizable songs in order to borrow from or comment upon the traditions they represent. This is incredibly important in improvisatory systems such as Jazz, where quoting Blues motifs or parts of others' solos are part of the conversational nature of the music and an important way musicians articulate both their power and respect for others in performance spaces. Quoting a musical canon signifies group membership by that performer, and the manner of performance can demonstrate complicated relationships to that canon, from respect to snarky derision. A great example of the latter is the quotation of *Moscow Nights* – the melody of which was used as a half-hour signal by Radio Moscow during the Cold War – by the infamous rock band Grazhdanskaya Oborona (GrOb) in their song "Кленовый лист" ("Maple Leaf").

There is obviously, then, a rich debate to be had over when a moment in music is meaningfully similar enough to another to be treated as the same in the minds of listeners. Empirical approaches could help define when musical ideas are considered perceptually similar to an arbitrary listener. In particular, it could be useful to quantify the notion of perceptual similarity, at least partially in terms of the features or components of given perceptual objects.

A Piece of Music

In order to attempt this practice with music, we ought to define the perceptual objects of sound, as well as their features. Any whole piece of music consists of many sounds, with particular events frequently containing *motifs*. A *motif* is a thematic element that ought to be recognized when repeated, and are frequently useful objects when analyzing musical works. My use of the term a *moment* in music is meant to include the occurrences of motifs or some discrete subset of a piece that is considered perceptually distinct.

Music is frequently conceptualized as the organization of sound over time, treating pitch and rhythm as its key features. These alone do not describe the full complexity of sound. *Timbre* may come to mind as a salient part of a listener's sonic experience that isn't captured by pitch or rhythm. Frequently defined as what allows a listener to distinguish between two sounding instruments that are otherwise producing sounds of the same pitch and loudness, timbre is difficult to parameterize simply (Tenney, 2015c). While this hasn't stopped exciting musical research into the physical components of timbre and our perception of different sounds (M. Lavengood, 2019; M. L. Lavengood, 2007), incorporating timbre into a model of perceptual similarity may be outside the scope of this project.

Were someone to sing, strum, or otherwise strike up some tune, they would produce a series of notes, to the delight of any listeners-on. The *pitch* of these notes could be described in terms of their frequency in Hertz, a measure of how often the air is displaced each second by the periodic wave resulting from their musical action. The higher this frequency, the higher in pitch the note will be perceived to be. Specifically, pitch height increases proportionally to the logarithmic increase in frequency. This means that sounds whose frequencies are 100 Hz apart are farther in perceived pitch at low frequencies than higher frequencies. Human listeners are able to perceive pitches in the range of 20 Hz to 20 kHz, though this range tends to decrease according to the natural hearing loss expected from age or other environmental factors.

When comparing multiple notes, examining the relationship between their frequencies can help identify the *interval* formed between them. For example, when hearing one note at 440 Hz and another at 880 Hz – when these frequencies form a 2:1 ratio listeners will hear two notes an octave apart. Many listeners would find this interval to be *consonant*, or aesthetically pleasing. According to Western musical notation, these would also both be the same kind of note, depicted as "A" notes at the interval of an octave. Different intervals, with different ratios to describe them, form scales that generate norms of harmonic and melodic practices. The perfect fifth (with a 3:2 ratio) and major third (5:4) compose the major triad, perhaps the cornerstone of Western tonality, familiar even to the ears of a nonmusician.[1]

Rhythm refers to how notes are organized in time. The majority of songs, especially those heard in Western societies, organize notes into four *beats* that regularly repeat, although other numbers of beats are not uncommon. Many dance musics from court traditions are organized into three beats (e.g., "Dance of the Flowers" from Tchaikovsky's ballet *The Nutcracker*), and songs organized into six beats are common in both classical and popular traditions (e.g., "Miss

[1] For a short example as evidence, go to https://youtu.be/JkFLF_k_XDk.

You" on Sound & Color by Alabama Shakes). Whatever the number of beats, these form a *pulse* or a grid underneath the whole song. Performers may play multiple notes in one beat, notes lasting several beats, or other subdivisions and syncopations, but will generally retain a perceptible pulse. This helps songs be danceable to an audience and more easily playable to a group of musicians.

Figure 1.

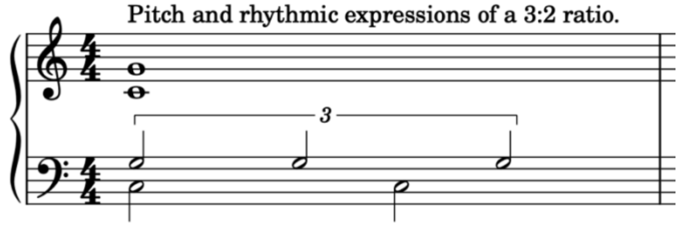
Excerpt from Tchaikovsky's Serenade for Strings in C, Op. 48, III.

In section A, beginning measure 45, the second violins (second line from the top) and violas (middle line) play harmony in triplets, a hemiola pattern, while the first violins (top line) and cello (fourth line from the top) sections play a melody with duple subdivisions.

Interestingly, our interval names for pitch intervals can also be applied to the rhythmic phenomenon of *polyrhythms*, where multiple streams of notes are played at different (coprime) pulses. That is, it's possible that in the time it takes one drummer to play the four beats one would expect in a disco song, a keyboard player could play three chords, forming a 4:3 polyrhythm. These grooves are more common in non-Western musical practices, especially in Afro-Cuban styles, but aren't impossible to come across in the works of Western composers. One example favored by the author can be found in the third movement of Tchaikovsky's *Serenade for Strings*. See Figure 1 for an excerpt of the piece featuring an extended 3:2 relationship between the instruments with harmonic and melodic roles. For an example of using intervals to describe both pitch interval and a polyrhythm, see Figure 2.

With these features in hand (or, if you will, "in ear"), we can attempt to determine what must be true of musical moments for them to be perceived as similar by a listener. First, though, I would like to emphasize that difference does not preclude similarity. Some moments in music may be perceived as similar, sharing some elusive quality that allows for experiential connections to emerge, even though they are different in terms of any combination of pitch, rhythm, timbre, or loudness, for example. Consider different performances of the United States National Anthem, a common case where the same song is played in different keys, by vocalists who wildly embellish the melody, and while being accompanied by all manners of instruments, depending on the setting. While the performances of Lady Gaga at President Biden's inauguration in January 2021 was *different* in many, many ways from that of Jimi Hendrix at

Figure 2.
Pitch and rhythmic expressions of a 3:2 ratio.



The interval of a perfect fifth between the C4 and G4 (top line) and the hemiola polyrythm (bottom line) can both be described by a 3:2 ratio, referring to pitch frequency and number of beats per time period of interest, respectively.

Woodstock in 1969, they may still be *similar* to listeners. This is also to point out that two musical moments could be trivially similar if they are perceived as the same, rather than different, on all dimensions. So, interesting claims about the similarity of musical moments will be found past the boundaries at which sounds are perceived as indistinguishably the same.

Boundaries for recognizing difference in sounds

We owe a knowledge of these boundaries at which we fail to recognize sounds as different to researchers in psychoacoustics and music cognition. While these literatures have frequently been more involved in matters of pitch, there are applicable insights into perceptual boundaries for the dimension of rhythm. Overwhelmingly, these findings are related to when pulses, or a series of even beats, are treated the same by performers. When building a computational model to account for exact onsets of notes played by musicians, (Large & Kolen, 1994) relied on the assumption that initial metric information determined a pulse grid that to-be-performed material would be fit to. This metric entrainment, as they refer to it, is highly related to other notions of oscillatory patterns and resonance systems discussed in theories of pitch perception, as we'll talk about shortly (Angelis et al., 2013; Large & Snyder, 2009; van Noorden & Moelants, 1999; Velasco & Large, 2011).

Understanding the overall pulse or meter, as well as the hierarchy of weak and strong beats, is useful when hoping to perform or analyze any piece of music. Notably, both intentional and unintentional departures from a strict pulse grid arise in musical performance. Intentional departures are frequently notated in music, and can provide great expressive power. Unintentional departures – real problems for researchers attempting to model why musicians play when they do – may reflect a combination of the kinesthetic difficulties of a musician producing a sound exactly when they would like to, as well as flexibility in how far from the grid can still constitute an appropriately timed note. Large and Kolen (1994) suggest that while it is hard to explain the exact misplacement of any note with respect to its nearest place in the pulse grid, in data provided by highly trained musicians tended these non-exact notes tended to be within a critical range of any given subdivision of the overall meter. While these subdivisions are also flexible within the context of a piece, this could imply that notes are metrically the same if they are attributable to the same place in a pulse grid.

Listeners' sensitivity to changes in tempo, or shifts to the underlying grid on which notes ought fall, vary according to the present tempo and whether the tempo is increasing or decreasing. The reported boundary is in terms of the just noticeable difference (JND), referring to how much the stimuli have to change in order for listeners to correctly report noticing a difference in more than 50% of cases. When asked to recognize a decrease in tempo (the music becoming slower), listeners reached above-chance accuracy when the tempo changed by around 6% of the initial tempo (in bpm). When asked to recognize an increase in tempo, the JND for these listeners was around 6% of the initial tempo at fast tempos (around 200 bpm) and increased to as much as 13% of the initial tempo at the slow tempo of 48 bpm (Dowling & Harwood, 1986; Lehman, 2012).

To remark on one complicating factor to this conclusion, note that different musicians and musicians of different practices place different amounts of attention to *where* within the expected range of a pulse grid notes tend to fall. That is to say, while notes are rarely perfectly placed on some beat, it does matter whether a musician is consistently placing notes ahead of or behind this beat. This artistic difference has been most clearly written about as a stylistic difference between Western classical music, which does aim to minimize individual and unintentional deviations from an ideal pulse grid, and improvisatory Jazz traditions that frequently employ different styles of swing or relationships to the beat in their playing (Ellis, 1991; Lehman, 2012). So, even empirically measurable sensitivity to changes in beat may vary according to musical training and exposure to different musical traditions, as the definition of what placement with respect to a grid is ideal is obviously dependent on these factors.

For the feature of pitch, we can describe the necessary physical difference between pitches such that a listener is able to correctly report their difference greater than half of the time. The JND for detecting differences in pitch varies according to other features of the sounds and by task demands. When notes are played in quick succession and listeners are asked to make a judgement about whether the sounds were the same or different pitch, they score above chance when the difference between the notes exceeds about 0.5% of the former's frequency (Justus & Bharucha, 2002), though this interacts with our logarithmic perception of pitch. The JND also varies according to the time between the pitches, with higher acuity for notes played simultaneously than consecutively, and for pure sine tones compared to notes with richer harmonic content (Borchert, 2011). Interestingly, human accuracy in terms of JNDs is not greater for pitch than features such as brightness or loudness, despite the greater musical weight given to pitch in most analytical and compositional practices (Cousineau et al., 2009, 2014; McDermott et al., 2010).

Most musical scenarios involve judging the relative size of intervals as they make up the contour of different melodies, or comparing these melodies themselves, rather than judging the similarity of two consecutive pitches. People with Western Classical musical training are able to be accurate in size judgments between two intervals when their size differed by as little as 100 cents, about the distance from one piano key to the next (if that piano, like most nowadays, is tuned in twelve-tone equal temperament), while nonmusicians are similarly accurate when the difference is slightly larger, at to 125 cents (Zarate et al., 2012). This difference according to musical training highlights the flexibility of this boundary according to learned musical structures. In fact, a sizable number of musical practices utilize differences in pitch smaller than a 12TET semitone of 100 cents. The difference between notes characterizing the particular *ragas* in Indian Classical music are as small as one twenty-secondth of an octave (approximately 55.54 cents), and those between *maqamat* in Arabic Classical traditions are in terms of quarter tones, or approximately 25 cents (Gann, 2019). Additionally, there exists a rich world of microtonal composers who through various techniques employ notes much closer in pitch than 100 cents (the "Hyperchromatica" collection by Kyle Gann makes for a fun entry point).[2]

In addition to ideas of mere proximity as a heuristic for the similarity of pitch ideas, a more complex notion of continuity may also guide our perception. Continuity, in my use, will refer to the influence of familiar musical systems on the perception of sounds. Work by Goldman et al. (2020) demonstrated that even among trained musicians, those who frequently improvise in musical practice show behavioral and neural differences when perceiving harmonic progressions whose second of three chords was sometimes varied. In fact, even mere exposure to different musical systems may be important in forming our perceptions of complex musical stimuli. Even in an experimental setting, when certain pitches are presented more frequently than others for a short period of time, people are faster to make recognition judgments and likely to rate a pitch as more pleasurable when presented with a more common pitch, compared to an uncommon one (Ben-Haim et al., 2014), and similar effects can be found when listeners are introduced to new, unfamiliar tonal systems (Sandbank, 2019).

[2] While the accompanying YouTube playlist includes "Rings of Saturn," you can also visit https://kylegann.com/Gannmusic.html for .mp3's of these and other pieces.

Western systems of tonality may interact powerfully with our perception of pitch intervals. When asked to make judgments about the size of the interval between two notes, between which a short musical example is played, accuracy in those judgments are higher when the musical example is tonal, or in a familiar key to a Western listener, rather than atonal (Graves & Oxenham, 2017). Neto et al. (2021) had students from Western conservatories in Canada and Brazil listen to a short primer, which could be either tonal (the A melodic minor scale) or atonal (an ascending set of non-repeating, unevenly spaced notes). After this, participants were played a set of two notes in A melodic minor forming either a minor third, major third, or perfect fourth, and asked to provide a subjective rating of the distance between the notes. While both minor and major thirds are two notes apart in the *scale* of A minor, minor thirds (three half steps wide, or 300 cents) are smaller than major thirds (four half steps, or 400 cents). By contrast, major thirds and diminished fourths are both intervals between two notes four half steps (400 cents) apart, but diminished fourths are three notes apart in the *scale*, and represent functionally more distant notes than a major third. When preceded by the *tonal* primer, participants rated the diminished fourths as larger than the major thirds, and those in turn larger than the minor thirds. When preceded by the *atonal* primer, the size difference between major thirds and diminished fourths disappeared, suggesting that these intervals are only perceived to be different sizes within a tonal context that classifies them as differently sized according to scale steps, at least among this sample of students attending music conservatories.

The harmonic series is a physical and theoretical system that may also highly influence our perception of the relationship between pitches. While the harmonic series has been significant to developments of Western music, both art and popular traditions, its influence may be distinct from that of the harmonic systems developed in Classical or Jazz practices, for example. (Demany & Ramos, 2005) played participants inharmonic chords consisting of sine tones at large, equal distances from each other (e.g., six sine tones each a major sixth apart). Following this, participants could either be played a note present in the preceding chord, one absent but about a half step (~100 cents) away from a note in the preceding chord, or one absent and about halfway between two notes in the preceding chord. While participants were accurate in reporting the presence of the present notes, and the absence of the half-step difference notes, participants tended to inaccurately report that the "halfway" target note *had* been present in the preceding chord. While by absolute proximity, these "halfway" target notes were more dissimilar to the previous chord than the target notes a half step away, the "halfway" targets seemed to be perceived as more consistent with the previous chord, at least enough to drive false positives in the recognition task. In order to explain the apparent difference in harmonic continuity participants attributed to these different kinds of target notes, researchers investigated the potential existence of frequency shift detectors. These hypothetical neural mechanisms are theorized to be attuned to small changes in absolute frequency between successive sounds, since these produce larger dissimilarities in two tones' harmonic series than larger changes in frequency (Demany et al., 2009; Demany & Semal, 2018).

In addition to harmonic schema, contour may also be a valid component of what makes a series of pitches continuous or not. Contour consists of directional information between subsequent pitches in a musical phrase, and can be visualized as the pattern of notes ascending, descending, or not moving. While when humans are asked to reproduce familiar melodies by singing them, they often do so in the same key as the original piece (Demany & Semal, 2007), familiar melodies can be recognized in any key since the exact intervals between notes are preserved through transposition. This is consistent with everyone's rendition of "Happy

Birthday" seeming to be in a different key than everyone else in the room; a melody can retain its identity regardless of the tonic center. Recognition of transposed melodies may not only depend on exact transposition, where all intervals are exactly preserved, though. While listeners seem to be able to distinguish melodies from musical phrases of the same length with random contour, they don't perform above chance when distinguishing exacting transpositions of melodies to ones with the same contour as the original (Dowling, 1978; Kleinsmith & Neill, 2018).

Grouping musical moments

The previous section details our understanding of when we can tell the difference between particular sound events according to their rhythm or pitch. This allows us to examine the interesting (non-trivial) cases where we may or may not find sounds to be meaningfully perceptually similar. However, the experimental settings relevant for determining our recognition for changes to the features of sounds include incredibly simplified and discretely delivered sounds. In answering questions about how we determine the similarity of musical moments more broadly, it's useful to find additional boundaries concerning how we group discrete sound events into musical moments at all. This section will outline current methods of understanding how we group successive sounds into related components of a common perceptual object, to the extent that we can explain musical moments in terms of musical features.

Looking first towards rhythm, a paper by London (2002) reviews psychoacoustic and psychological investigations of metric perception, including that of *subjective rhythmization,* or when we perceive subsequent notes to be forming beats. On the fast end, we stop perceiving these beats when the inter-onset interval (IOI) between notes exceeds around 100 ms, analogous to a measure of notes at a tempo of 600 beats per minute (bpm). There exist a few *metric envelope*s, or regions of time in which we tend to group hierarchical information. These have musical significance, since it would be possible to hear a measure containing six notes either as six independent notes, as three sets of two, as two sets of three, and sometimes as one full beat containing six notes. Contextual information interacts with these metric tendencies to inform what groupings we hear.

The fastest of these metric envelopes is when notes have IOIs of 200-250 ms, corresponding inversely to a measure in 240-300 bpm. At this speed, subdivisions are rare and would tend to be simple rather than compound (splitting beats into two rather than three components), both for the sake of performers' physical capacities and for the perception of listeners. The second metric envelope overlaps strongly with the range in which people are most comfortable spontaneously creating a pulse – when asked to tap at a comfortable and even speed, for example. Beats are most strongly felt with IOIs of 600-700 ms, or at a tempo of 85-100 bpm. At the slowest end, notes with IOIs of 1500-200 ms, or at a tempo of 20-40 bpm, form a lower limit at which we are comfortable grouping notes in one pulse. Interestingly, this tends to be a highly subdivided meter, so that listeners hear pulses at lower hierarchical significance at the reportedly more comfortable level around 600 ms, for example. While musical practice overwhelmingly tends to align with these regions, pieces such as John Cage's "ASLSP" (As Slow as Possible) – currently 20 years into its 639-year performance – push these practical and perceptual boundaries in the name of artistic experimentation.

Research in auditory scene analysis investigates whether listeners explicitly report hearing audio as either one or two "streams" of audio – that is, whether diotically presented sounds are perceived as a single unit, or two separate ones. Evidence from this field is consistent with proximity being an important principle in how and whether we associate sounds. Work by Snyder et al. (2008) showed that when participants heard a repeated pattern of two notes, they were more likely to report hearing two distinct "streams" or sources of sound as the interval between the two notes increased. When the notes were an octave apart, greater than 95% of listeners reported hearing these notes as separate streams. In addition to this effect of the interval between notes on a given trial, participants' perception of either one or two streams was also significantly affected by the intervals they heard in previous trials, even as long as 15 seconds later. Having heard the notes in unison in a previous trial increased the likelihood participants would hear two streams in the current trial when hearing any interval greater than a unison, with the reverse effect for having previously heard an interval of an octave.

This anchoring effect, where previously heard sounds seem to change the parameters of expected sounds in the future, is consistent with previously discussed literature describing the effect that musical systems such as tonal systems and the harmonic series have on perception. Additionally, composers have employed processes of time-dependence in generative compositional processes. Markov chains have been one way of computing the likelihood of a subsequent note given features of the previous note. As one example, the *Illiac Suite* (1957) algorithmically determined the intervals between notes based on judgements of the proximity between notes as well as their harmonic relationship or continuity (Ames, 1989).

James Tenney, a music theorist and composer, has written extensively about methods of algorithmically determining how sections of music are likely split into smaller perceptual units (Tenney, 2015b, 2015a). His goal has been to make rigorous the definition that when one unit of music is more internally similar than similar to neighboring units, this drives perceptual cohesion of the similar unit, and distinction from other nearby units. Decisions about these groupings are made by integrating information about multiple features of music, including time, pitch class, and the intensity of the notes. Tenney's models depend on weighting each of these features, so that they are linearly combined to compute holistic similarity between musical moments.

With the coding help of Larry Polansky, their mathematical model analyzed the compositions of a few composers, producing sketches of the perceptual objects at different hierarchical levels. The weights for each musical feature found to be ideal varied according to composer, and are summarized in Figure 3. Tenney noted the difference in the weights for the parameter of intensity representing a tendency for the markings of fortissimo or pianissimo dynamics, for example, to be structural rather than expressive decisions for Varèse and Webern compared to Debussy. However, Tenney remarks that the weights for pitch were mostly arrived at through trial and error, with no clear theoretical – or statistical – rationale governing the selection process.

Figure 3.

Features and weights reported by Tenney (1978).

	duration	*pitch*	*intensity*	*timbre*
Varèse	1.0	0.67	6.0	20.0
Webern	1.0	0.5	6.0	0.0
Debussy	1.0	1.5	2.0	0.0

Duration referred to 10 ms segments of time, pitch to the number of half steps between notes, intensity (loudness) to an ordinal difference in notated dynamics (e.g, there is one unit between a mezzoforte and forte), and with timbre referring to dummy variables corresponding to each instrument in a given piece.

This model divided pieces into moments at different hierarchical levels, divisions that were useful for further music theoretical analysis conducted by Tenney of the selected pieces. Certainly, assuming the model is effective in dividing the score into units similar to those perceived in the mind of a listener, it is useful to base score analyses on these divisions rather than ones based purely on reading written music, or even through a dialogic process of re-listening and re-marking an understanding of the piece in written form. However, there are several drawbacks to the model as proposed and worked through in 1978. Tenney's model computes several levels of hierarchical groupings, with each higher-level grouping computed in succession; first, all of the smallest units ("elements") are identified, and then the model runs through the whole piece again to group these “elements” into “clangs,” and so on up to the level of the whole piece. This iterative process by the model is likely a departure from the human ability to perceive different groupings at different hierarchies simultaneously when listening to a piece of music.

Another notable difference in the model’s computational process arises from the operationalization of the feature of pitch. Tenney notes that pitch was computed by the number of half steps between two adjacent notes (the absolute difference in their ascribed MIDI number). This computation is therefore done without respect to the scalar role of notes, information which we know from work by Neto et al. (2021) affects the perception of interval size. Additionally, this computation doesn’t take into account the surrounding harmonic context accompanying any pairs of notes; we know from work by many scholars in music cognition that the harmonic context can affect recognition judgements, and any music theorist or performer would tell you that the harmonic function of a set of notes changes depending on their association with other chordal material present. So, Tenney’s model may still lack the power to incorporate harmonic information into its division of pieces into perceptual units.

Tenney writes about another significant drawback of his work at that point: these different features, while weighted differently, are still linearly combined to define the holistic similarity or dissimilarity between subsequent notes. Already, we have found ways in which tonality and rhythm interact to determine whether or not listeners can recognize differences in notes on axes of pitch and rhythm (e.g., E. M. O. Borchert, 2011; Graves & Oxenham, 2017). How and whether we group musical moments into perceptual objects also seems to depend on interactions between these features. For example, in a study by Moelants and van Noorden (2005), participants were played looped polyrhythms that varied in different aspects of pitch and rhythm (overall tempo, polyrhythm density, and pitch interval), and told to tap along to the beat however the saw fit. When the pitch interval was greater – as the two notes in the polyrhythm were farther apart – participants were less likely to tap in time to the overall beat. They instead tapped in time to one of the two notes in the polyrhythm, but this pattern seemed to depend on the polyrhythm and overall tempo. When the polyrhythm was less dense (e.g., 5:2 rather than 5:4), people were more likely to tap along to the fast component rather than the slow component. Similarly, at slower tempos, participants were more likely to tap in time to the fast component of the polyrhythm. Prince et al. (2009) found that the delay between the musical example and the onset of the second note affected judgments about the interval between the first and second note under certain circumstances. When the musical example was tonal, the delay did not affect accuracy; however, when the musical example was atonal, accuracy was significantly higher if the second note was played on the beat established by the example, rather than off the beat.

Summary

From this multidisciplinary approach to our perception of sound, we have gained many useful frameworks to guide our inquiry into judgements of perceptual similarity between musical moments. Beyond the boundaries at which we can ascribe difference to musical sounds, we know that different sounds can be considered similar if they are in agreement with each other in terms of familiar musical systems to listeners. Notions of resonance and harmonic series relationships govern metric entrainment and many harmonic systems of pitch; training in certain musical practices and short-term exposure to certain sounds can change which heuristics of continuity are most salient when judging musical material; features such as pitch and rhythm interact in nonlinear ways when we form holistic perceptual judgements of musical moments.

There remains ample room for further research to understand more precisely how certain types of listeners ascribe similarity or difference to musical moments, and under what conditions different or competing heuristics from familiar musical systems are employed to guide these judgements. However, I would like to address one key inference that poses a weakness in the work we have reviewed so far concerning musical perception: we have not been measuring perception. Psychological research has depended on measuring behavioral responses, such as when participants can report recognition or when and how they produce sounds by singing or tapping. Music theoretical work has depended on a dialogic engagement between physically denotable divisions of a piece and an interactive representation of the musical work in the analysts' mind; Tenney and Polansky tuned the weights of their model so that the divisions produced by the computations were in line with their ideas about where perceptual objects should be in the pieces.

In order to strengthen the body of literature investigating music perception, I hope to find a way to make inferences about people's perceptual experiences without relying on their behavioral responses to music directly. As one way of forward, I will borrow from the psychological study of memory.

Context-Dependent Memory

An active subset of memory research focuses on context-dependent memory (CDM), a theory which states that when someone learns target material in a given *context*, they will do better on a test of that material when that context is present, rather than absent, during the test. Conceptually, the definition of context can refer to anything and everything that is not the target material itself: features of the room someone is in, how hungry they are, their mood and wandering thoughts, the sound of people talking a room away, the smell of paper in front of them, the din of computers buzzing behind them, and so on. Experimentally, researchers focus on manipulating features of a background environment that are complex but temporally stable, so that these contexts are associated with a longer event rather than a small moment within a lab procedure (Stark et al., 2018). To be functionally useful, different contexts must also exceed perceptual thresholds to be considered different, and must have some degree of behavioral relevance; while a context doesn't need to be explicitly presented as related to the target task (and many studies do not direct participants' attention directly to the context), if a context is not salient enough to enter at least pre-conscious awareness, it will not be an accessible part of memories formed during the task. The theory of context-dependent memory emerges from our understanding of episodic memory, and is related to the *encoding specificity principle*, which states that a memory for learned information or events includes not only the target information, but other information present during encoding such as task demands, how the material was learned, and other extraneous detail. Evidence for context-dependent memory has been found in a diverse range of such contexts, including but not limited to odor (Ball et al., 2010; Cann & Ross, 1989), state (Eich, 1980), incidental room environments (for review, see Smith & Vela, 2001) and imagined rooms (Masicampo & Sahakyan, 2014), and – of special interest to the present study – background music (Balch et al., 1992; Balch & Lewis, 1996; Isarida T. K. et al., 2008; T. K. Isarida et al., 2017; Mead & Ball, 2007; S. M. Smith, 1985).

Computational models of memory and neurological research have worked jointly to refine theoretical and practical motivations in the study of context and memory. The temporal context model (TCM) sought to provide a unified explanation of the recency and contiguity effects seen in free recall (Howard & Kahana, 2002). The recency effect refers to more recently presented material being more likely to be recalled than older material, while the contiguity effect refers to an asymmetrical effect where words presented close together in time are more likely to be recalled together, and such that words are most likely to be recalled in the same order as they were learned (i.e., recalled in sequential order rather than backwards). The components of the model include a slowly drifting representation of temporal context that is bound to a representation of items during encoding, with later updates to the memory of this item involving joint representation of the previous and current temporal context. Polyn et al. (2009) expanded TCM to detail a model of context maintenance and retrieval (CMR) that accounts not only for temporal context, but for list context and inter-item associations such as words' semantic connections, providing additional explanatory power for source and semantic clustering effects in free recall paradigms.

Neurological evidence for this slowly drifting temporal context has been found in electrocorticographic recordings of the temporal lobe and in whole-brain analyses (Manning et al., 2011). A body of animal studies involving lesions to the hippocampus have found such lesions to inhibit animals' ability to respond to previously learned contextual information, and their ability to respond appropriately to changes in contextual information (D. M. Smith et al., 2004), and the hippocampus is taken to be critical for integrating contextual and target information in episodic memory. Additionally, during recall processes Manning et al. (2011) found evidence for the reinstatement of context while retrieving target information. The hippocampus has also been recorded as sending information critical to distinguishing different periods of a task (e.g., earlier or later during a learning phase) to regions such as the cingulate cortex and anterior thalamic regions (S. M. Smith, 2009), and these as well as regions like the dorsolateral prefrontal cortex may play an important role in updating memory representations with previously integrated contextual information (Polyn & Kahana, 2008).

Theoretical explanations of context-dependent memory have emphasized the existence of multiple, complex components of the overall context present during encoding. For example, the *mental context hypothesis* states that the overall context during learning includes details about one's environment, their mood, thoughts, and associations with learned material (S. M. Smith, 1995). While it follows that memory for learned material is likely to be better when one specific part of the learning context – for example, a happy mood – is also present during a test rather than absent, the mental context hypothesis also accounts for reasons this may not be the case. Since one's mood is not the only component of these contexts, it may not also be an important enough component on any given task to drive a context-dependent memory effect. Even when the maintenance of mood context is enough to contribute to improved memory, forgetting due to changes in other aspects of context – the temperature of a room, for example – may still occur. In fact, someone's representation of the slow drift of time and of the type of task they're performing may be inextricable changes to someone's mental context between learning and a test.

One initial corollary to this hypothesis is that memory is more greatly affected by multiple changes to context between learning and test periods, compared to more simple changes. For research in place-dependent memory (for review, see Smith & Vela, 2001), evidence for this included greater effect sizes for CDM effects when context manipulation included changing the room environment, experimenter identity, and different internal factors for a participant, compared to only manipulating the room environment (T. Isarida & Isarida, 2014). This may be in part due to only certain manipulations crossing some threshold necessary for changes to be significant under task demands. In terms of the mental context hypothesis, given a greater proportion of dissimilarity between one's overall mental contexts at learning versus during a test, we can expect fewer context cues to be readily available to facilitate recall, resulting in poorer memory performance.

This has made it especially compelling when manipulating only one aspect of context produces a reliable effect, such as one of the earliest CDM studies that assigned participants to either an underwater or above-water context between learning and test (Godden & Baddeley, 1975), manipulating specific odors in the same room environment (Mead & Ball, 2007), or manipulating the tempo or key of a musical piece independently (Balch & Lewis, 1996; Mead & Ball, 2007). Of course, these manipulations don't preclude other aspects of mental context functioning as covariates; to this end, there has been a significant effort towards teasing out the contribution of mood, in particular, toward these effects (Balch & Lewis, 1996; Eich & Metcalfe, 1989; T. K. Isarida et al., 2017).

Task demands often influence the boundaries and importance of different contexts. For example, work found an effect of mood-dependent memory for words generated by participants in a given mood state, but not for words decided and presented by experimenters (Eich & Metcalfe, 1989). A hypothesis of mood's mediation of place-dependent memory effects, first proposed by (Eich, 1995), has been weakened by some evidence of moderating factors such as this; if mood as a context does not drive memory effects in the robust set of situations in which place-dependent memory effects have been found, it's hard to build a case that mood is a unique mediating factor for such effects. This is corroborated by evidence from animal brain studies, which have found that differences in patterns of hippocampal neuron firing are produced not only when the geometry of a room environment change, but also when task demands, perceived autonomy, or the types of rewards offered are manipulated (S. M. Smith, 2009).

Music-Dependent Memory

In the first study to establish background music as a context that could elicit context-dependent memory benefits, Smith (1985) found that participants who were tested on previously studied words after a 48 hour delay showed decreased levels of forgetting if they listened to the same background music at test as they had while studying the words. In Experiment 1, participants either heard Mozart's Concerto No. 24 in C minor, "People Make the World Go Around" from Milt Jackson's Sunflower jazz record, or no sound while studying a list of common words and during an initial, immediate test of how many words they could recall. When participants returned two days later, they were administered a surprise delayed test of free recall while either the Mozart, jazz piece, or no sound played in the background. For those who studied with music in the background, the number of words recalled at the delayed test was higher if they listened to the same selection, rather than the different selection or no sound. For those who studied with no music in the background, their ability to recall words during the delayed test was not significantly changed by whether music was played at the delayed test, providing some evidence against suggestions that memory effects are more caused by the distraction of background music. Experiment 2 replicated this general finding, and also showed that white noise was able to similarly function as a background noise context.

Subsequent work on music-dependent memory focused on teasing apart what features of background music may be most important for eliciting the CDM benefit. Work by Balch et al. (1992) used four different instrumental pieces that varied in genre (either Classical or Jazz) and in tempo (either slow or fast), and found that the proportion of words participants could recall during a surprise test was most disrupted when they heard music with a different tempo at test, compared to those who heard music of a different genre at test. Replicating this tempo-dependent memory effect, Balch and Lewis (1996) found that there was a stronger CDM benefit for tempo compared to genre in Experiment 1, and compared to timbre in Experiment 2. This work controlled for the key of included pieces, with the first author playing all selections in C major. In addition to this effect of tempo, Mead and Ball (2007) demonstrated that manipulating the tonality of a piece could produce a CDM effect, using Chopin's Waltz in A Minor, either played in the minor key as written or in A Major by a professional pianist.

Work by Isarida et al. (2017) challenged the strength of these findings in a similar study, where participants learned words while hearing a piece of music that was either fast or slow, and either in a major or minor key. Performance in a surprise test of recall was greater for those who heard the same piece of music rather than a different one during the final test, replicating the general CDM effect. However, those who heard a different piece of music at test did not perform significantly different from each other whether the piece of music had the same or different tempo or tonality to the original piece heard during learning. That is, the similarity of two different musical pieces' features such as tempo, tonality, or so on is not always sufficient to ameliorate memory detriments expected when the background music context is different. The authors' exclusion of a condition where participants heard the same piece of music that varied in tempo or tonality at test makes it difficult to draw conclusions about whether manipulating those features is sufficient to make the altered musical piece be perceived as a different piece, resulting in weakened memory performance.

Methodological Review

The broader literature of CDM research has a fair amount of methodological variance, with effects found in recognition as well as recall tests, when studying words as well as visual information such as faces, using indirect measures of memory, and differences in the delay between learning and test, just to name a few. The subset of studies focused on background music as a manipulable context are fewer and in some ways more consistent. My goal here is to note in what ways the present study was consistent with this literature, and where it departed.

All cited music-dependent memory studies test verbal memory rather than visual memory. Interestingly, a small study by Echaide et al. (2019) demonstrated that instrumental background music affected initial and future recall of visuospatial items, but did not impact similarly measured memory for words, suggesting that the use of words as target information rather than images is more useful if researchers hope to find music-dependent memory effects. These studies also almost always present words visually and test participants on them in a written format, although Smith (1985)'s Experiment 2 provided some evidence that the CDM effect is more pronounced when words are presented aurally rather than visually. However, aural presentation of words is not common in other music-dependent memory studies, and poses technological difficulties when researchers don't have fine control over how participants listen to audio. While studies investigating contexts such as odor have found significant effects for tests of recognition (Ball et al., 2010; Cann & Ross, 1989), and Smith and Vela (2001)'s meta-analysis found evidence for context effects when testing recognition, studies of music-dependent memory have exclusively utilized tests of free recall. Therefore, the present study tested memory for words presented visually through a delayed free recall test.

Though Smith (1985) exclusively measured recall after a 48 hour delay, Balch et al. (1992) replicated a music-dependent memory effect for an immediate test of recall but could not find a consistent effect for the test after a 48 hour delay. Subsequent work consistently utilizes immediate tests of recall, often after a relatively short delay (ranging from 30 seconds to 5 minutes). During these delays, many authors played intentionally distracting music (work by

Balch and colleagues (1992, 1996) featured atonal bamboo flute music, and Mead and Ball (2007) favored birdsong) in order to reduce potential effects of distraction for those who heard different or altered music compared to those who heard the same piece during the test. So, the present study follows suit, employing a relatively immediate test of free recall after a delay shorter than five minutes. During the delay period, participants will listen to pink noise while performing a visuospatial task. While pink noise is not likely to be as distracting as the sounds used in previous studies, it provides some control over the auditory context of participants during this phase, so that the transition to the test phase is comparable between participants; additionally, the manipulations of musical stimuli in the present study concern specific, small changes to harmonic information, so the delay period sound was selected to not contain confounding harmonic or melodic information.

Another point of difference between the Smith (1985) study and others is the exact mechanism by which participants learned words. Smith (1985) had participants study words intentionally, for an expected immediate test of free recall. They then found an effect of musical context on a surprise test of free recall after a 48 hour delay. Utilizing intentional learning is beneficial for the non-associative processing it may encourage in participants while studying, and Smith and Vela (2001)'s meta-analysis found that for incidental room environments, the mean weighted effect size for CDM effects were significantly lower when the processing of words at encoding was associative (d = .13), rather than non-associative (d = .33) or otherwise not specified (d = .38). However, work by Isarida et al. (2008) found no effect of musical context on participants' memory of words studied intentionally, when a test of free recall was employed after a 30 second delay. These authors did find a significant effect of musical context on participants' memory of words studied incidentally, where participants were shown each target word individually and asked to audibly state as many verbal associates as possible in the five seconds for which the word was presented. This is consistent with many other recent studies in music-dependent memory, which utilize incidental learning, short delays between learning and test, and have produced significant effects of musical, genre, and tempo as contexts (Balch et al., 1992; Balch & Lewis, 1996; Isarida et al., 2017; Mead & Ball, 2007). In order to more closely replicate the methodologies of more recent work in music-dependent memory and employ a paradigm in which there already exists evidence that context changes can differentially affect recall, the present study employed incidental learning, with participants rating a subset of the words used by Mead and Ball (2007) for the pleasantness on a likert-type scale.

Related to the manner of learning, there is conflicting evidence on whether the number of times words are shown to participants affects the strength of context-dependent memory effects. Within participants who learned words incidentally, Isarida et al. (2008) found in a within-subjects comparison that there was an effect of context on the recall of words presented once, but not those presented twice during the learning phase. The authors concluded that presenting words twice strengthened their representation while diminishing their association with the surrounding background-music context. However, Mead and Ball (2007) did find significant effects of background music's key on participants' free recall of words presented twice in random order. While it is unclear what produced the null effect in Isarida et al. (2008)'s study but not that of Mead and Ball (2007), the present study is more methodologically similar to the latter than the former: I used English rather than Japanese words, had participants rate words for pleasantness rather than verbally report associates, and did not manipulate the number of times words were presented within subjects. Therefore, I opted to display words twice in a random order during the incidental learning phase.

Perhaps even more significantly, to the best of my knowledge every study investigating music-dependent memory has thus far operationalized music as background music, such that their stimuli consisted of rich, complex instrumental musical examples, overwhelmingly pulled from Western Classical repertoire, occasionally also featuring American Jazz pieces. This is not to say that these pieces have been employed without rigor. While most studies justified their selection of pieces as ones likely not to be familiar to their college-aged participants, Mead and Ball (2007) also reported results of a pilot study that verified that students at their institution tended to rate the chosen Chopin waltz as "neither particularly familiar nor unfamiliar" (12). Additionally, although most studies used a single musical selection per condition (e.g., one piece that was both slow and in a minor key, one piece that was fast and in a major key, etc.), authors Isarida and Isarida frequently employed multiple selections per condition, in order to present results that could more robustly be explained by shared features of these pieces rather than particularities of single examples (Isarida T. K. et al., 2008; T. K. Isarida et al., 2017).

Even using verifiably unfamiliar musical selections and varying specifical musical features while controlling for target ones, it is not far-fetched to say that there remain similarities and dissimilarities not controlled for between selected pieces: the timbre of instruments, melodic contours, arrangement techniques, chord progressions, harmonic or melodic structure, differential salience of an instrument in a given moment, the overall mood or social context invoked by a piece of music, may all vary in ways uncontrollable and sometimes inarticulable. All of these musical features may connect moments of music in surprising ways, and may evoke other memories in surprising ways. Effect sizes of context change on memory are greater when multiple features of a context are changed at once, compared to when single features are changed, a common finding that mental context theory offers explanation for. So, efforts to report significant effects of background music as a context are strengthened when changes in background music are complex. The present study is an effort to begin inquiry into what combinations of musical features produce these rich and sometimes deeply personal subjective experiences of music, and connections between musics. The effect sizes of hypothesized effects are likely to be smaller than other studies, but differences would be strong evidence for music-dependent memory effects dependent specifically on the feature of pitch interval – a perceptual feature that is complex in its own right.

Present Study

The present study investigated whether the framework of a context-dependent memory experiment would be a valid way to assess the perceptual categorization of musical examples by varying the pitch interval of a simple piece of background music between its presentation during the learning and the test of words. This study focused on only manipulating the dimension of pitch of one note, in order to alter the interval formed between two pitches. Out of a desire to maintain some complexity to rhythm in order to retain some generalizability to other musical situations, these notes were complex pitches played back in a polyrhythm. In particular, I chose a 5:4 polyrhythm played at 150 BPM, which Moelants and van Noorden (2005) found to be a combination at which tapping preferences between the fast-versus-slow components and the high-versus-low notes to be split most evenly.

Key to the motivation of this study is the notion that context-dependent memory effects are driven by the complex and multiple components of a given context. The integration of contextual information and target material is in part built on associations between the features of the context – here, the timbre of notes, perceived rhythmic emphasis, and pitch content – and the features of the target material being studied – here, English words. Differences in the pitch interval of the musical context may be critical for some of these context-target associations. Additionally, a context may be usefully recalled to aid retrieval based on its holistic representation, distinct from merely the sum of its features. Here, while changes to absolute pitch distance may be a significant change to this feature of the musical context, if the impression of the harmonic material is not severely altered, physically different sounds may still cue the same global impression of the original musical context.

So, the present study sought first to replicate an expected music-dependent memory benefit, testing the effect of hearing either no audio, or the polyrhythm at a same or a different pitch interval, on delayed recall of learned words. It was expected that those who heard the same audio would recall more words than those who heard no audio at test. If the difference of pitch interval is sufficiently perceptually distinct, those who heard different audio would be expected to recall fewer words than those who heard the same audio. Furthermore, if pitch interval is of unique importance to the present background sound context, those who heard different audio would be expected to not perform differently than those who heard no audio during the delayed test of recall. However, if remaining similarities between the different audio and the original audio are still beneficial to contextual reinstatement processes, those who heard different audio would be expected to recall more words than those who heard no audio at test.

While this comparison can give insight into how crucial the broad construct of pitch interval is to a musical context, further investigation is necessary to tease out how pitch interval creates musical contexts. Specifically, the current study classified pitch intervals according to the *octave level* and *interval class* of the interval created between the two notes of the polyrhythm. Octave level categorically defines the register of the interval, such that an interval smaller than one octave is in the first octave level, but one between two and three octaves is in the third octave level. The interval class of an interval refers to the music theoretical name of an interval, irrespective of octave displacement – i.e., we will call the distance between a C and a G a perfect fifth, no matter how many octaves are between the particular notes C and G. Compared to those who heard the same audio – with both the same interval class and octave level – at test as during learning, I hypothesized that fewer words would be recalled by those (1) who heard the polyrhythm at a greater octave level at test, and (2) who heard different interval classes than that of the perfect fifth heard during the learning phase.

Pilot I

It is well-known that different musical intervals can evoke different subjective experiences in listeners. Therefore, a small pilot study was conducted to inform the selection of sounds for the main experiment. Out of the 12 interval classes, four were selected that (a) each did not offer significantly different subjective experiences at different octave levels, (b) did not significantly differ from each other in these subjective experiences, and (c) satisfied musical constraints.

Method

Participants

Participants were recruited either through social media or through Amazon's Mechanical Turk (MTurk) with the use of TurkPrime by CloudResearch (Litman et al., 2017) between December 28, 2020 and February 11, 2021. Participants recruited through social media received compensation of $6.25 for an approximately 30-minute task, in accordance with New York State minimum wage as of January 1, 2021; participants recruited from MTurk received compensation of $3.75 for the task after providing a valid completion code, in accordance with the United States Federal minimum wage.

While 72 participants completed some portion of the task and had data stored on Inquisit's servers, this included only 61 complete responses (M_{age} = 31.82, SD_{age} = 12.88). Of these, participants were excluded who failed to identify the direction between pairs of notes more than half the time, who could identify the correct musical name for intervals all of the time, who reported turning off audio during the task, and/or who reported not providing intentional answers during the task. After these measures, 45 participants (M_{age} = 33.51, SD_{age} = 13.95) were included in analyses. For full demographic information, see Table 1.

Table 1. Demographic Data in Pilot Experiment

	Full Data	Analyzed Data
	n	*n*
Race		
African-American/Black	2	2
Asian	7	5
Hispanic/Latino	4	3
White/Caucasian	51	38
sum	64.00	48.00
Gender		
Female	32	22
Male	20	15
Nonbinary	8	7
Other	1	1
sum	61.00	45.00
Platform		
and	1	1
ios	6	3
mac	20	15
win	34	26
sum	61.00	45.00
Musical Training		
No	23	15
Yes	38	30
sum	61.00	45.00
Plays Music		
No	37	26
Yes	24	19
sum	61.00	45.00

	Full Data		Analyzed Data	
	mean	*sd*	*mean*	*sd*
Age	31.82	(12.88)	33.51	(13.95)
Years Musical Training	2.00	(1.62)	1.50	(0.71)
Weekly Playing Hours	17.39	(22.98)	12.78	(15.12)
Weekly Improv Hours	4.46	(4.03)	3.67	(2.59)
Weekly Music Listening	10.65	(12.09)	10.86	(13.39)

Materials

Musical Stimuli. All musical stimuli consisted of a 5:4 polyrhythm played at pitch intervals ranging from a minor second to three octaves apart, for a total of 36 possible pitch intervals. These form three distinct octave levels, and thirteen possible interval classes, the name used to refer to an interval (e.g., a perfect fifth and a major third are different interval classes). Stimuli were created in Musescore (Schweer, 2020) and exported to .mp3 files to be played through the Inquisit Web 6 player (2020). At each pitch interval, stimuli were looped indefinitely with a period of 1600 ms, equivalent to a tempo of 150 beats per minute (bpm), to equalize the perceptual salience of each rhythmic component (Moelants & van Noorden, 2005). For full description of musical stimuli, see Appendix A.

Interval Recognition Task. In order to confirm participant's self-reports about hearing ability, perfect pitch or pitch blindness, and functioning audio equipment, they completed a one-minute interval recognition task in the Inquisit Web 6 player. Participants were asked to make judgments about the direction and quality of six intervals, played both melodically (so that participants heard the first and then the second note) and harmonically (both notes played at once). Notes were played as quarter notes at 150 bpm; the audio example lasted 5 s. After hearing the interval, participants were first asked whether the second note was higher or lower than the first note, and then were asked to either select which musical interval name best reflected the interval they heard, or respond "I do not know". Participants who provided the incorrect direction for four or more of the six intervals, or who provided the corrected quality for all six intervals, had their data removed from main analyses.

Lexical Decision Task. Participants completed an adapted version of the Lexical Decision Task available in the online Millisecond Test Library (K. Borchert, 2020) in the Inquisit Web 6 player. Lists of words and nonwords were generated through the English Lexicon Project (Balota et al., 2007), selected to be comparable in length and such that the English words were high in concreteness and neutral in valence. The full list of words is available in Appendix B. During the approximately three minute task, a word or nonword was presented on screen briefly (250 ms), followed by a fixation cross for the duration of the response period (700 ms). During the response period, participants used key presses to categorize the characters as either a word or a nonword. Participants were instructed to respond as fast as possible while maintaining accuracy. The accuracy and reaction time, measured from stimulus onset, of their judgments were recorded.

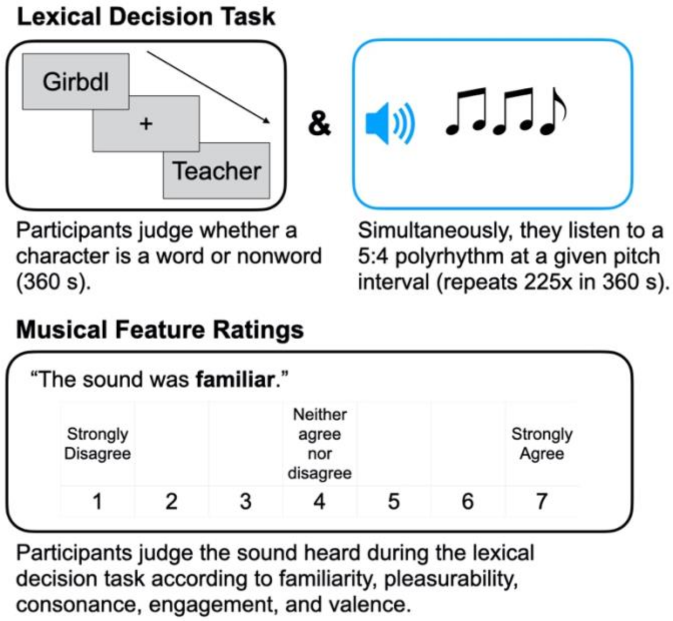
Musical Feature Ratings. Participants were asked to make judgments about the musical features of sounds they heard using likert-type scales in the Inquisit Web 6 platform. They were presented with statements such as “This sound was familiar” and were asked to indicate their personal agreement with the statement on a scale of 1 (strongly disagree) to 7 (strongly agree), with 4 as a neutral midpoint (neither agree nor disagree). For each sound, they judged familiarity, pleasure, consonance, engagement, distraction, and valence.

Demographics and Musical Experience Questionnaires. Participants were asked to provide their age in years, gender identity, and any applicable race or ethnicity labels in demographics questionnaires. Additionally, they were asked to provide information about musical training and experience, instrument practice habits, and music listening habits.

Procedure

Participants who were determined to qualify and provided informed consent for the present study completed the experiment in the Inquisit Web 6 player (*Inquisit 6 Web*, 2020). They first completed the interval recognition task, confirming audio playback on their device. For the main task, participants heard all 12 possible interval classes in a random order, and each interval class at one of three octave levels. See Figure 4 for an overview of the main task.

Figure 4.
Procedure for the pilot experiment.



For each pitch interval at which participants heard the musical stimuli, they completed both the lexical decision task and the musical feature ratings. Each sound would play on a repeated loop for the duration of the lexical decision task and during the musical feature ratings, for an approximate total of two minutes. After completing this for all pitch intervals, participants were provided with demographic and musical experience questionnaires. Finally, they were thanked, debriefed, and provided payment.

Results

Measures

In order to adjust for the repeated measures of participants' accuracy, reaction time, and responses to the musical feature questions, z-scores were computed for these eight dependent variables for each of the 12 intervals a given participant heard, with respect to the participant's mean for the variable across all pitch intervals. For example, if a participant rated the perfect fifth as more "happy" than the average of all twelve "happy" ratings they provided, the z-score for the perfect fifth's happy rating would be some positive number. A measure of overall abnormality was computed for each of the 36 pitch intervals heard by participants by taking the root sum squared of the z-scores corresponding to reaction time and the six musical feature questions. A targeted measure of abnormality was pre-registered to be computed in a similar manner as overall abnormality, including only the musical feature questions whose z-scores were significant predictors of z-scores for reaction time or accuracy. However, two simultaneous regressions found that no standardized musical features significantly predicted standardized reaction time or standardized accuracy (in both models, all *p*'s > 0.05 for predictors, and both $R^2_{adj} < 0.01$, *p*'s > 0.05). Therefore, the targeted measure of abnormality was not computed; when it would have been used, analyses were conducted on the eight dependent variables separately.

Pitch Interval Selection

The first pre-registered criteria for sounds to be selected for experimental use is that the overall abnormality associated with an interval class should not significantly vary across its three octave levels. To test whether this within-interval-class variation occurred, a grouped one-way ANOVA was conducted, analyzing variance in overall abnormality by octave level, grouping by interval class. Because this consisted of 12 simultaneous tests, $a = 4.166\text{x}10^{-3} = (0.05/12)$ was taken as a Bonferroni-adjusted threshold for significance (Jafari & Ansari-Pour, 2019); at this level, no interval classes significantly varied in overall abnormality by octave level.

The interval classes of the perfect fourth and fifth were of special interest for this question: since one of them would serve as the learning phase interval for all experimental participants, these should especially not vary in overall abnormality by octave level, or other selected intervals should vary similarly by octave level. The perfect fourth did not vary by octave level ($F(2,42) = 0.353$, $p = 0.704$). However, the perfect fifth trended toward varying by octave level ($F(2,42) = 2.986$, $p = 0.061$), with post-hoc analyses using Tukey HSD pairwise comparisons showing that this was due to the perfect fifth at the first octave level being rated as less abnormal than usual ($M = 1.396$, $SD = 0.950$) compared to at the second ($M = 2.017$, $SD = 0.741$) and third octave levels ($M = 2.110$, $SD = 0.699$).

In order to determine whether any pitch intervals differed from others in overall abnormality, a 3x12 ANOVA was conducted with the factors octave level and interval class. This found a significant effect of interval class, $F(11,504) = 1.517$, $p = 0.019$. However, a Tukey HSD post-hoc test found no significant pairwise differences after correcting for multiple tests, all p's > 0.05. Therefore, further analyses were conducted to see whether the octave level and/or interval class had an effect on the eight standardized dependent variables: reaction time, accuracy, and the six musical feature ratings. Because this consisted of conducting eight 3x12 ANOVAs simultaneously, $a = 6.25x10^{-3} = (0.05/8)$ was taken as the Bonferroni-adjusted threshold for statistical significance. At this level, significant effects of pitch interval features were found for four out of six musical features, but not for standardized reaction time or accuracy.

For standardized ratings of pleasure, there were significant effects of octave level ($F(2,504) = 7.34$, $p = 7.21x10^{-4}$), interval class ($F(11,504) = 6.786$, $p = 1.12x10^{-10}$), and their interaction ($F(22,504) = 1.646$, $p = 3.3x10^{-3}$). For standardized ratings consonance, there were significant effects of octave level ($F(2,504) = 10.502$, $p = 3.4x10^{-5}$), interval class ($F(11,504) = 5.502$, $p = 2.51x10^{-8}$), and their interaction ($F(22,504) = 1.857$, $p = 1.1x10^{-3}$). For standardized ratings of engagement, there were significant effects For standardized ratings of distraction, there were significant effects of octave level ($F(2,504) = 10.083$, $p = 5.09x10^{-5}$) and interval class ($F(11,504) = 2.896$, $p = 0.001$). Finally, for standardized ratings of happiness, there were significant effects of octave level ($F(2,504) = 3.206$, $p = 4.1x10^{-3}$) and interval class ($F(11,504) = 12.881$, $p = 2.09x10^{-23}$).

To further investigate these effects, four Tukey HSD post-hoc tests were conducted for standardized scores of pleasure, consonance, distraction, and happiness, with $a = 0.0125 =$

(0.05/4) as the Bonferroni-adjusted threshold for statistical significance. Participants rated sounds at the third octave level as less pleasurable than usual (*M* = -0.186, *SD* = 0.802) compared to sounds at the second (*M* = 0.102, *SD* = 0.898, $p_{adj} = 2.4x10^{-3}$) or first octave level (*M* = 0.087, *SD* = 0.949, $p_{adj} = 5.66x10^{-3}$). Similarly, participants rated sounds at the third octave level as less consonant than usual (*M* = -0.225, *SD* = 0.802) compared to sounds at the second (*M* = 0.130, *SD* = 0.882, $p_{adj} = 1.61x10^{-4}$) or first octave level (*M* = 0.099, *SD* = 0.985, $p_{adj} = 9.23x10^{-4}$). Finally, participants rated sounds at the third octave level as more distracting than usual (*M* = 0.237, *SD* = 0.875) compared to sounds at the second (*M* = -0.150, *SD* = 0.959, $p_{adj} = 1.29x10^{-4}$) or first octave level (*M* = -0.091, *SD* = 0.932, $p_{adj} = 2.06x10x10^{-3}$). See Table 2 for a visualization of these comparisons.

Table 2.

Significant differences in interval class according to selected musical features.

	1	2	3	4	5	6	7	8	9	10	11	12
1. Minor second		B: †; D: ***		A: ***; B: ***; D: ***	A: **; D: ***; B: **		A: **; D: ***		A: *; D: ***	D: **		A: **; B: ***; D: ***
2. Major second						D: **		D: *				
3. Minor third				D: *						C: *		
4. Major third						A: ***; B: **; D: ***		A: **; D: ***		B: *	A: *; B: *; D: **	
5. Perfect fourth						B: *; D: ***	A: *	D: ***	A: *			
6. Tritone							A: *; D: ***		D: ***	D: *		A: *; B: *; D: ***
7. Perfect fifth								D: ***			D: *	
8. Minor sixth									D: **			D: ***
9. Major sixth												
10. Minor seventh											C: *	
11. Major seventh												B: *
12. Perfect octave												

Note : Significance levels: *: $p < 0.0125$, **: $p < 0.00125$, ***: $p < 0.000125$. A: pleasure; B: consonance; C: distraction; D: happiness. Empty cells denote no significant pairwise difference. Cells containing † were only significant between specific octave levels of the given interval classes.

Moderation Analyses

Exploratory analyses were performed to investigate the potential moderating effect of various features of the participant pool, including from which online source participants were recruited, their variance in musical experience, and whether participants reported altering the volume of sounds on their devices.

A simultaneous regression tested whether the factors online source (Qualtrics, MTurk, or unsure[3]), musical training (yes or no), and current musical playing (yes or no) could significantly account for variance in the overall abnormality of participants' subjective experience with these sounds. The model as a whole accounted for a small but significantly greater than zero amount of variance in overall abnormality (R^2 = 0.018, p = 0.009), and only found participants being sourced from Qualtrics to be a significant predictor of overall abnormality, b = 0.248, p = 0.014.

Some participants reported altering the volume of audio playback at some point during the task, and were not excluded from the primary analyses. An independent samples t-test evaluated differences in overall abnormality scores for the twelve pitch intervals rated by given participants, finding no significant difference in these scores between participants who did or did not report altering volume (t(538) = -1.317, p = 0.188). Furthermore, repeating the analyses relevant to pitch interval selection having excluded participants who reported altering volume did not alter the direction of any results, and did not produce new pairwise conflicts between interval classes on any of the musical feature ratings.

---

[3] Participants who did not successfully submit their Inquisit completion code on the payment confirmation pages of Inquisit or MTurk could not be linked to their source.

Discussion

Pitch Interval Selection

Out of the 12 interval classes included in this study, four will be selected to be included in a further experiment. Either the perfect fourth or perfect fifth will be included as the interval class heard at the first octave level during the learning phase of a context-dependent memory paradigm. The included intervals should not vary or should vary similarly in participants' subjective experiences, and should not vary or should vary similarly by octave level. Musical theoretical considerations provide further constraints: no two of the four included intervals should be musical inversions of each other, there should be a balance of consonant and dissonant intervals, and an ideal set of intervals would be balanced in the difference between interval sizes.

The perfect fourth was selected over the perfect fifth to be included in the further experiment, due to the trend toward within-interval-class variance observed with all participants. Although this finding is not robust, it is important that the sound to be heard during the learning phase of the context-dependent memory paradigm does not vary significantly by octave level if any conclusions are to be drawn about the manipulation of octave level independently of interval class in analyzing the experiment's results.

The post-hoc analyses of musical features according to interval class and octave level provided further insight into which intervals created dissimilar subjective experiences to participants. Since included interval classes should not significantly differ from each other on these metrics, and the perfect fourth was to be included, I first analyzed which interval classes were significantly different from the perfect fourth. The minor second and tritone differed from the perfect fourth in their ratings of pleasure, consonance, and happiness; the minor sixth additionally differed from the perfect fourth in their ratings of pleasure and consonance. Between the remaining seven interval classes (this does not include the perfect fifth, the musical inversion of the perfect fourth) and the perfect fourth, there were nine sets of four intervals that contained neither musical inversions nor interval classes found to be different on any musical feature ratings, summarized in Table 3.

Table 3.

Characteristics of viable options for pitch interval selection.

Option 1:	Major second	*+1*	Minor third	*+2*	**Perfect fourth**	*+7*	Perfect octave
Option 2:	Major second	*+2*	Major third	*+1*	**Perfect fourth**	*+4*	Major sixth
Option 3:	Major second	*+2*	Major third	*+1*	**Perfect fourth**	*+7*	Perfect octave
Option 4:	Major second	*+3*	**Perfect fourth**	*+4*	Major sixth	*+2*	Major seventh
Option 5:	Major second	*+3*	**Perfect fourth**	*+4*	Major sixth	*+3*	Perfect octave
Option 6:	Minor third	*+2*	**Perfect fourth**	*+5*	Minor seventh	*+2*	Perfect octave
Option 7:	Major third	*+1*	**Perfect fourth**	*+4*	Major sixth	*+3*	Perfect octave
Option 8:	**Perfect fourth**	*+4*	Major sixth	*+1*	Minor seventh	*+1*	Major seventh
Option 9:	**Perfect fourth**	*+4*	Major sixth	*+1*	Minor seventh	*+2*	Perfect octave

Potential sets of four pitch intervals are listed above with distance in half steps between consecutive intervals.

Notice that options 4-6 are the only ones without intervals a half step apart in size, and without intervals greater than a tritone (six half steps) apart in size. Of these, notice that in terms of general consonance, option four contains only dissonant intervals in addition to the consonant perfect fourth. On the other hand, option five contains three intervals including the perfect fourth that could be termed consonant – the perfect fourth, major sixth, and perfect octave. So, it's the sixth option that provides the best balance of consonant and dissonant intervals. Additionally, the minor seventh plays a special role of being an experientially dissonant note, but harmonically consonant with the perfect fourth, with their higher notes being a perfect fourth themselves. So, the set containing the minor third, perfect fourth, minor seventh, and perfect octave was selected based on these data.

Effects of Pitch Interval

In this pilot study, I measured the effect of pitch interval both on lexical decision task performance and on participants' subjective experiences. Performance on the lexical decision task, measured both by standardized accuracy and standardized reaction time, a) was not significantly predicted at a given pitch interval by standardized musical feature ratings, and b) never varied according to features of pitch interval overall. Therefore, there seems to be an important distinction between a participant's cognitive ability and their subjective experience of different sounds, with cognitive ability as measured by task performance not being significantly affected by differences in pitch interval.

Subjective experiences, on the other hand, varied a fair amount. No significant differences in pitch interval were found for familiarity or engagement, while there were significant differences for features carrying some aesthetic or emotional valence: pleasure, consonance, distraction, and happiness. This is consistent with explanations of different intervals as primarily having different emotional and sensational qualities. For example, the intervals that in Table 2 can be seen to have been rated as significantly different from the minor second in terms of happiness included almost the entirety of the major scale (with one exception: the minor second was significantly different from the minor, rather than major, seventh). That is, not only were aesthetic and emotional features worlds in which sounds were found to differ, but they differed in ways consistent with musicological ideas of differences between intervals.

In addition to musical theory with respect to interval class, remember that with greater distance between pitches, we can expect the pitches to be less harmonically and melodically associated with each other. An interesting set of findings in support of this were the main effects of octave level on standardized musical feature ratings, where compared to participants' average ratings, sounds at the third octave level were rated as less pleasurable, consonant, and happy – but more distracting – than sounds at the first and second octave level. Additionally, the pairwise comparisons investigating the interaction between interval class and octave level found no individual cases where a sound at the third octave level was significantly *more* pleasurable, for example, than another sound at the first octave level. This could suggest that these more distant sounds are less often evaluated or able to be evaluated along the same emotional or aesthetic axes that listeners would usually employ.

Evaluating how differently participants rated a given pitch interval on a feature like distraction compared to how distracting they usually found sounds (using the standardized measures of musical features) is invaluable for the project of avoiding future use of pitch intervals that drive particularly abnormal subjective experiences for participants. In order to fully contextualize these experiences, it is useful to additionally observe the raw, non-standardized ratings of pitch intervals according to different musical features, as seen in Figure 5. Of primary concern are the ratings for distraction: with remarkable consistency, participants rated sounds as maximally distracting.

Pilot II

The previous pilot found concerningly high ratings of distraction for sounds presented to participants.[4] Accordingly, this pilot study tested whether refined musical stimuli, with a more naturalistic timbre, could counteract the levels of distraction and unpleasantness experienced by participants, and further inform the selection of four interval classes to be used in the experimental portion of this study.

[4] The pre-registration for the second pilot was submitted before the discovery of an error in the computation of z-scores for distraction, an error discovered and corrected after the initial analyses for the second pilot. This did not affect the high raw scores for rating in either pilot, but did affect the computation and normality of overall abnormality and standardized ratings of distraction. While the pre-registration for the second pilot noted irregular z-scores and a need for nonparametric analyses, neither of these issues persisted after the computation of z-scores for distraction was adjusted.

Method

Participants

Participants were recruited through Amazon's Mechanical Turk (MTurk) with the use of TurkPrime by CloudResearch (Litman et al., 2017) between February 26 and February 27, 2021. Participants received compensation of $3.75 for the task after providing a valid completion code, in accordance with the United States Federal minimum wage. While 63 participants completed some portion of the task and had data stored on Inquisit's servers, this included only 41 complete responses (M_{age} = 44.15, SD_{age} = 12.88). Of these, participants were excluded who failed to identify the direction between pairs of notes more than half the time, who could identify the correct musical name for intervals all of the time, and/or

Table 4. Demographic Data in Pilot Experiment

	Full Data	Analyzed Data
	n	*n*
Race		
Asian	6	4
Hispanic	2	2
White	33	25
sum	41.00	31.00
Gender		
Female	21	15
Male	20	16
sum	41.00	31.00
Platform		
mac	3	2
win	38	29
sum	41.00	31.00
Musical Training		
No	29	21
Yes	12	10
sum	41.00	31.00
Plays Music		
No	37	27
Yes	4	4
sum	41.00	31.00

	Full Data		Analyzed Data	
	mean	*sd*	*mean*	*sd*
Age	44.15	(12.88)	45.84	(13.07)
Weekly Music Listening	8.34	(6.56)	8.81	(6.19)

those who reported turning off audio or not providing intentional answers during the task. After these measures, 31 participants had their data included in analyses (M_{age} = 45.84, SD_{age} = 13.07). For full demographic information, see Table 4.

Materials and Procedure

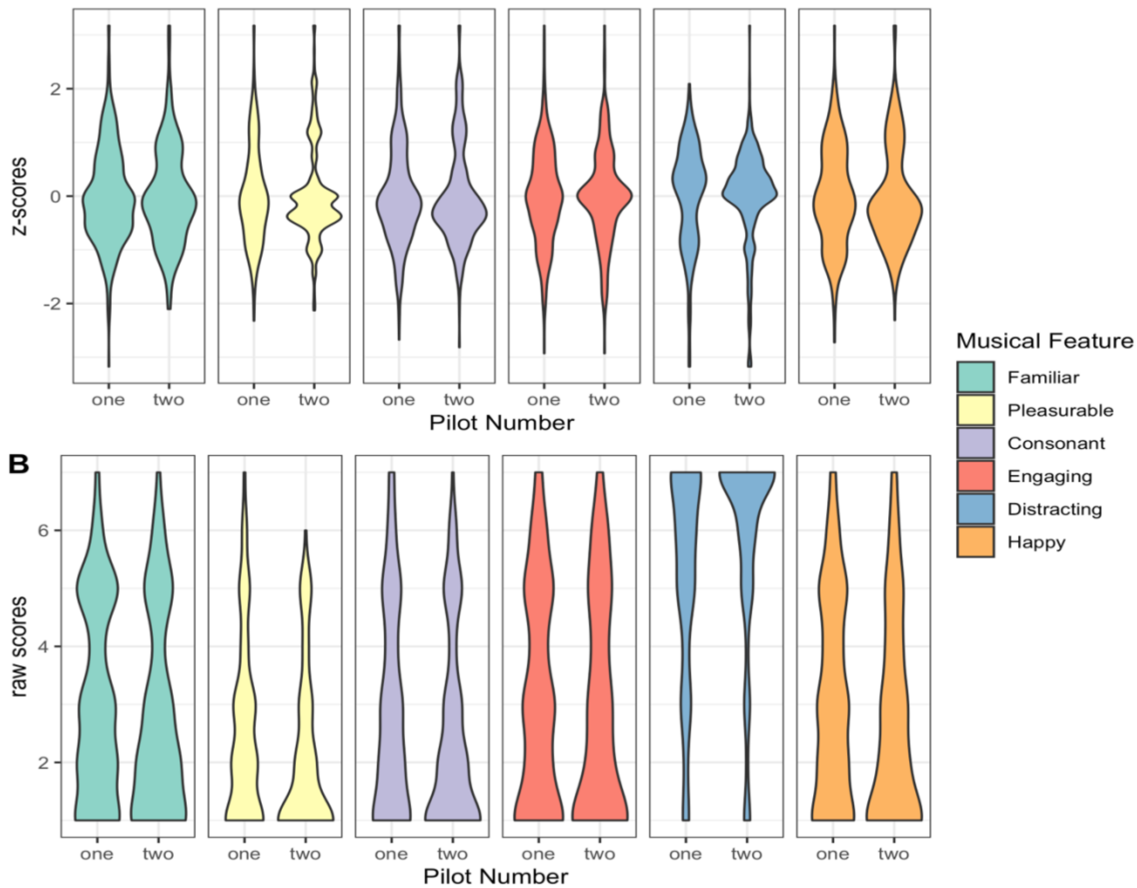
The design of Pilot II was identical to Pilot I except for the production of musical stimuli. The sounds used both in the interval rating task, and the 5:4 polyrhythms at 36 pitch intervals were created in Musescore and played on the “Mellow Steinway” from a soundfont developed by John Nebauer and published under a creative commons license.

Results

Comparing Musical Stimuli

In order to investigate whether the musical stimuli changed in pilot two elicited different subjective experiences in participants, a grouped independent samples t-test was conducted on raw scores for each of the six musical features between pilot one and two, with $a = 8.3\text{x}10^{-3}$ as the Bonferroni-adjusted threshold for significant differences. This found significant differences between the mean ratings of four musical features: pleasure ($t(910) = 4.456$, $p_{adj} = 9.41\text{x}10^{-6}$), consonance ($t(910) = 3.812$, $p_{adj} = 1.47\text{x}10^{-4}$), distraction ($t(910) = -5.35$, $p_{adj} = 1.11\text{x}10^{-7}$), and happiness ($t(910) = 3.198$, $p_{adj} = 1.43\text{x}10^{-3}$), with ratings for familiarity ($t(910) = 2.59$, $p_{adj} = 9.75\text{x}10^{-3}$) and engagement ($t(910) = 1.674$, $p_{adj} = 9.45\text{x}10^{-2}$) trending toward significance. As visible in Figure 5, while ratings for distraction tended to be higher in the second pilot compared to the first, ratings for all other features tended to be lower in the second pilot. For all of these effects, neither the direction nor significance level were affected when comparing the second pilot to a random subset of the first pilot matched in size, or when only comparing results from participants recruited through MTurk.

Figure 5. Normalized and raw musical feature ratings between pilots



Violin plot showing central tendency and density of z-scores (top) and raw ratings (bottom) for musical features.

Pitch Interval Selection

Since the additional goal of the second pilot was to evaluate whether altered musical stimuli also altered the interval selection process, similar analyses were conducted to evaluate whether standardized ratings of musical features, as well as reaction time and accuracy, were different between the two pilots. A grouped independent t-test found no significant differences between any z-score for dependent variables according to pilot, all p's > 0.95. A separate independent t-test found no significant difference between overall abnormality between the pilots, $t(910) = 1.177$, $p = 0.239$. As such, analyses to select interval classes for future use should be comparable between the first and second pilot; to evaluate this, these analyses are repeated on only the data from the second pilot, and on both datasets simultaneously.

Analyzing the Second Pilot. The first pre-registered criteria for sounds to be selected for experimental use is that the overall abnormality associated with an interval class should not significantly vary across its three octave levels. To test whether this within-interval-class variation occurred, a grouped one-way ANOVA was conducted, analyzing variance in overall abnormality by octave level. Since this was grouped by interval class, $a = 4.16\text{x}10^{-3}$ was taken as the Bonferroni-adjusted threshold for significance; at this level, the effect of octave level was not significant. However, as before, the intervals of the perfect fourth and perfect fifth were inspected individually. This found that within the pitch interval of the perfect fourth, overall abnormality trended towards varying by octave level, $F(2,28) = 4.854$, $p = 0.015$. A Tukey HSD post-hoc test found that overall abnormality for trials where participants heard the perfect fourth was lower when it was played at the third octave level ($M = 1.4$, $SD = 0.633$) compared to the second octave level ($M = 2.55$, $SD = 0.819$, $p_{\text{adj}} = 0.0125$). The perfect fifth did not trend towards varying by octave level, $F(2,28) = 1.279$, $p = 0.294$.

In order to determine whether any pitch intervals differed from others in overall abnormality, a 3x12 ANOVA was conducted with the factors octave level and interval class. This found a significant effect of octave level, $F(2,336) = 4.161$, $p = 0.016$. Pairwise comparisons using a Tukey HSD post-hoc test found that overall abnormality was lower for sounds heard at the third octave level ($M = 1.729$, $SD = 0.811$) compared to both the second ($M = 2.036$, $SD = 1.06$, $p_{adj} = 0.033$) and first octave levels ($M = 2.083$, $SD = 0.931$, $p_{adj} < 0.01$).

Further analyses were conducted to see whether the octave level and/or interval class had an effect on the eight standardized dependent variables: reaction time, accuracy, and the six musical feature ratings, with *a* = 6.25×10^{-3} as the Bonferroni-corrected threshold for statistical significance. At this level, no significant effects were found for octave level or interval class on any of the dependent variables.

Analyzing Both Pilots Simultaneously. The first pre-registered criteria for sounds to be selected for experimental use is that the overall abnormality associated with an interval class should not significantly vary across its three octave levels. To test whether this within-interval-class variation occurred, a grouped one-way ANOVA was conducted, analyzing variance in overall abnormality by octave level, grouping by interval class, with *a* = 4.16×10^{-3} as the Bonferroni-adjusted threshold for significance. This did not find cases where the effect of octave level was significant. While the perfect fourth trended toward varying by octave level ($F(2,73) = 2.620$, $p = 0.080$), the perfect fifth did not ($F(2,73) = 0.910$, $p = 0.407$).

In order to determine whether any pitch intervals differed from others in overall abnormality, a 3x12 ANOVA was conducted with the factors octave level and interval class. This found a significant effect of interval class ($F(11,876) = 2.682$, $p = 0.002$) and octave level ($F(2,876) = 4.850$, $p = 0.008$), but not their interaction ($p = 0.285$). Post-hoc analyses were conducted using Tukey HSD pairwise comparisons only on the main effects. These found that overall abnormality was lower for sounds at the third octave level ($M = 1.863467$, $SD = 0.8261392$) compared to the second ($M = 2.07$, $SD = 0.981$, $p_{adj} = 0.016$) and first ($M = 2.067$, $SD = 0.966$, $p_{adj} = 0.018$) octave levels. Additionally, overall abnormality for the minor seventh ($M = 1.716$, $SD = 0.848$) was lower than at the perfect octave ($M = 2.289$, $SD = 1.063$, $p_{adj} = 0.007$), but higher than the minor second ($M = 2.219$, $SD = 1.046$, $p_{adj} = 0.023$) and major third ($M = 2.218$, $SD = 0.963$, $p_{adj} = 0.021$).

Further analyses were conducted to see whether the octave level and/or interval class had an effect on the eight standardized dependent variables: reaction time, accuracy, and the six musical feature ratings, with *a* = 6.25×10^{-3} as the Bonferroni-corrected threshold for statistical significance. These revealed significant effects for all six musical features, but none for reaction time or accuracy.

For standardized ratings of familiarity, there was a main effect of both octave level ($F(2,876) = 6.026$, $p = 0.003$) and interval class ($F(11,876) = 2.737$, $p = 0.002$). For standardized ratings of pleasure, there was a main effect of both octave level ($F(2,876) = 12.929$, $p = 2.93 \times 10^{-6}$) and interval class ($F(11,876) = 9.857$, $p = 5.1 \times 10^{-17}$). For standardized ratings of consonance, there was a main effect of both octave level ($F(2,876) = 11.966$, $p = 7.46 \times 10^{-6}$) and interval class ($F(11,876) = 7.464$, $p = 2.43 \times 10^{-12}$), as well as their interaction ($F(22,876) = 2.04$, $p = 3 \times 10^{-3}$). For standardized ratings of engagement, there was a main effect of interval class ($F(11,876) = 3.658$, $p = 4.28 \times 10^{-5}$). For standardized ratings of distraction, there was a main effect of octave level ($F(2,876) = 11.292$, $p = 1.44 \times 10^{-5}$). For standardized ratings of happiness, there was a main effect of both octave level ($F(2,876) = 5.103$, $p = 6.0 \times 10^{-3}$) and interval class ($F(11,876) = 13.724$, $p = 1.64 \times 10^{-24}$).

To further investigate these effects, Tukey HSD post-hoc tests were conducted for all six musical features, with *a* = 8.33×10^{-3} as the Bonferroni-corrected threshold for statistical significance. Table 5 displays the significant pairwise differences by interval class and the interaction between interval class and octave level. Ratings of familiarity were higher than normal for sounds at the second octave level ($M = 0.08597038$, $SD = 0.9048883$) compared to the third ($M = -0.13816525$, $SD = 0.8202827$, $p_{adj} = 4.46 \times 10^{-3}$), though no comparisons with the first octave level were significant ($M = 0.04751817$, $SD = 0.9148821$). Ratings of pleasure were lower than usual for sounds at the third octave level (*M* = -0.195, *SD* = 0.758) compared to those at the second (*M* = 0.095, *SD* = 0.919, p_{adj} = 4.85x10^{-5}) or first (*M* = 0.094, *SD* = 0.918, p_{adj} = 6.25x10^{-5}). Ratings of consonance were lower than usual for sounds at the third octave level (*M* = -0.197, *SD* = 0.789) compared to those at the second (*M* = 0.11, *SD* = 0.924, p_{adj} = 2.8x10^{-5}) or first (*M* = 0.081, *SD* = 0.965, p_{adj} = 3.03x10^{-4}). Ratings of distraction were higher than usual for sounds at the third octave level (*M* = 0.205, *SD* = 0.788) compared to those at the second (*M* = -0.109, *SD* = 0.939, p_{adj} = 4.9x10^{-5}) or first (*M* = -0.089, *SD* = 0.948, p_{adj} = 1.82x10^{-4}). Finally, ratings of happiness were higher than usual for sounds at the second octave level (*M* = 0.111, *SD* = 0.955) compared to the first *M* = -0.098, *SD* = 0.911, p_{adj} = 0.007), though no comparisons with the third octave level were significant (*M* = -0.017, *SD* = 0.880).

Table 5.

Significant differences in interval class according to selected musical features for both pilots.

	1	2	3	4	5	6	7	8	9	10	11	12
1. Minor second		E: *		B: ***; C: ***; D: *; E: ***	B: ***; C: *; E: ***		B: ***; E: ***		E: ***	E: *		B: *; C: ***; E: ***
2. Major second				A: *; B: ***; C: **; E: *								
3. Minor third				B: ***; C: * C: ***; E: ***			E: **					D: *
4. Major third					C: †	B: ***; C: ***; E: ***	C: *	B: ***; C: **; E: ***	C: *	B: **; C: ***; E: *	B: ***; C: ***; E: ***	
5. Perfect fourth						B: **; E: ***		B: *; E: *	C: †			
6. Tritone							B: ***; E: ***		E: ***			C: **; E: ***
7. Perfect fifth								B: **; E: ***			B: *	
8. Minor sixth												E: *
9. Major sixth												C: †
10. Minor seventh												C: *
11. Major seventh												C: *
12. Perfect octave												

Note : Significance levels: *: $p < 0.0125$, **: $p < 0.00125$, ***: $p < 0.000125$. A: familiarity; B: pleasure; C: consonance; D: engagement; E: happiness. Empty cells denote no significant pairwise difference. Cells containing † were only significant between specific octave levels of the given interval classes.

Discussion

Pitch Interval Selection

The first point of difference between the analyses from the first pilot study and those incorporating the results from the second pilot came from investigating whether certain interval classes varied in overall abnormality by octave level. Given the Bonferroni-adjusted alpha level, no interval classes significantly varied according to octave level in either study. However, marginal trends between the first and second pilot varied, with the perfect fifth trending towards varying in the first pilot, and the perfect fourth in the second. Neither of these are robust findings, although it's notable that the trend towards variation for the perfect fourth was replicated when analyzing the data combined between the pilots, despite the smaller sample size of the included data from the second pilot. Using similar logic to that in the discussion of the first pilot, the replication of the trend to variance within the interval class of the perfect fourth suggests that the perfect fifth should be selected instead of its inversion, the perfect fourth.

Other pitch intervals included alongside those with the interval class of the perfect fifth should not be significantly different from the perfect fifth or each other in the subjective experiences reported by participants, and no two selected interval classes should be musical inversions. Based on the pairwise comparisons displayed in Table 5, there were five interval classes that could be selected in addition to the perfect fifth: the major second, major sixth, minor seventh, and perfect octave. Note that the major second and minor seventh are musical inversions of each other, so they could not both be selected. This left two potential sets of four interval classes: either (a) the major second, perfect fifth, major sixth, and perfect octave, or (b) the perfect fifth, major sixth, minor seventh, and perfect octave. Notably, both of these selections could also have been possible selections given the pairwise comparisons using only data from the first pilot study.

Musical considerations informed the decision between these two options. The primary difference between the sets involved the size differences between the different intervals. In the first set, there are five, two, and three half step size differences between subsequent pairs of intervals. This is pleasing, and alludes to an additional useful relationship within this set of intervals: they are producible by stacking perfect fifths above a tonic, forming the perfect fifth first, followed by (an octave and) the major second, followed by (an octave and) the major sixth. The ratings of these intervals on each of the six musical features are summarized in Table 6.

Musical Stimuli

While standardized ratings of musical features and standardized performance on the lexical decision task did not differ significantly between the pilots, raw ratings of musical features did differ, with ratings in the second pilot tending to be less familiar, pleasurable, consonant, engaging, and happy, but more distracting than the first pilot, on average. This remained true even when evaluating a subsample of the first pilot's data to control for the difference in sample size between the studies. While this was unexpected given the refinement to the musical stimuli included in the second pilot, it is important to note that the refinement was

Table 6. Characteristics of selected pitch intervals.

	Familiarity		Pleasure		Consonance		Engagement		Distraction		Happiness	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Major second	2.68	(1.62)	2.33	(1.50)	2.64	(1.70)	3.16	(1.98)	5.25	(2.01)	2.97	(1.80)
Perfect fourth	2.99	(1.72)	2.64	(1.56)	3.05	(1.82)	3.18	(1.80)	5.30	(1.95)	3.28	(1.68)
Major sixth	2.99	(1.83)	2.42	(1.61)	2.70	(1.71)	3.17	(1.87)	5.50	(1.87)	3.25	(1.83)
Perfect octave	3.03	(1.80)	2.78	(1.94)	3.28	(2.07)	3.33	(2.09)	5.11	(1.98)	3.51	(2.04)

Raw scores were provided in a 1-7 scale, with 1 indicating a low amount of the feature, and 7 a high amount.

not dramatic, and that no participants were asked to explicitly compare the two sounds. It remains possible that time confounds (the first pilot being completed by MTurk participants on Monday, February 1 and Wednesday-Thursday, February 10-11, while the second pilot was completed by participants on Friday-Saturday, February 26-27) or other uncontrolled features between the two studies contributed to the differences in raw ratings. For example, responses to an open question soliciting feedback at the end of the task frequently included remarks about the length of exposure to the sounds affecting the overall experience of the task – since both pilot studies exposed participants to 12 pitch intervals, this aspect of the study remained unchanged and may have contributed to raw ratings of musical features. Usefully, while participants in pilot two tended to rate sounds in more unfavorable ways, the difference in their ratings across different pitch intervals were not systematically different. This suggests that the between-pitch-interval subjective experiences were comparable across pilots, which makes sense: while timbral changes were made to the stimuli, the pitch content was not affected.

The hypothesis that this timbral improvement would benefit raw scores for musical features of sounds was not supported by the second pilot's data. However, the combination of data from both pilots supported the selection of sounds that included the pitch interval of the perfect fifth. In the data from the first pilot alone, the perfect fifth varied in overall abnormality by octave level, a feature not conducive to selection. Using the refined musical stimuli independently and in combining this data with that using the original stimuli, this within-interval-class variation was found for the perfect fourth, but not the perfect fifth. As a result, experimenter judgement was such that the set of intervals supported by combined data and data from the second pilot would be used in the following experiment, and would be played using the refined musical stimuli, which supported the use of the perfect fifth.

Experiment

The main experiment evaluated the effect of manipulating a 5:4 polyrhythm's pitch interval on memory performance in a surprise test of recall. Participants who heard the same sound (at the interval class of a perfect fifth and at the first octave level) were expected to recall more words than those who heard a different or no sound, replicating general context-dependent memory effects. Of novel interest was the effect of hearing different interval classes (either the major second, perfect fifth, major sixth, or perfect octave) and different octave levels (the first, second, or third). Participants who heard the perfect fifth were expected to recall more words than those who heard sounds at a different interval class, and those who heard a sound at the first octave level were expected to recall more words than those who heard sounds at greater distances.

Method

Participants

Participants (N = 285; M_{age} = 40.33, SD_{age} = 12.45) were solicited through Amazon Mechanical Turk (MTurk), were U.S. residents, and were determined to have no hearing abnormalities and to consider English a primary language through a separate screening questionnaire, for which all participants were compensated $0.25. For their completion of the approximately 15-20 minute main task, participants were paid $2.50. The task was completed in the Inquisit 6 Web player.

While there were 388 unique, completed responses to the task, participants were excluded who reported either or both turning off audio during the task or providing disingenuous answers during the task, and those who failed to identify the direction between pairs of notes in the interval recognition task more than half the time or who could identify the correct musical name for intervals all of the time. See Table 7 for demographic information for the full and included participants.

Materials

Musical stimuli. All musical stimuli consisted of a 5:4 polyrhythm played at twelve different pitch intervals, selected to be comparable across listener responses based on results from the pilot study. The same audio files used in the second pilot study were used in the present experiment These pitch intervals are divisible into three different octave levels and four different interval classes: the major second, perfect fifth, major sixth, and perfect octave. For full notation, see Appendix A.

Interval Recognition Task. In order to confirm participant's self-reports about hearing ability, perfect pitch or pitch blindness, and functioning audio equipment, they completed a one-minute interval recognition task, as described in

Table 7. Demographic Information for Experiment Participants

	Analyzed Data	Full Data
	n	*n*
Race		
Asian	22	33
Black	14	28
Hispanic	12	23
Native	3	4
White	245	320
sum	296.00	408.00
Gender		
Female	148	205
Male	136	181
Nonbinary	1	2
sum	285.00	388.00
Platform		
and	22	29
ios	7	23
mac	36	48
win	220	288
sum	285.00	388.00
Audio Equipment		
headphones	121	178
speakers	164	210
sum	285.00	388.00
Musical Training		
No	146	224
Yes	139	164
sum	285.00	388.00
Plays Music		
No	235	329
Yes	50	59
sum	285.00	388.00

	Analyzed Data		Full Data	
question	*mean*	*sd*	*mean*	*sd*
Age	40.33	(12.45)	39.29	(12.52)
Music Training	8.37	(9.83)	8.16	(9.72)
Weekly Improv Playing	2.42	(1.73)	2.64	(2.64)
Weekly Music Listening	9.49	(9.73)	9.81	(9.69)
Weekly Music Playing	3.11	(2.58)	3.44	(3.34)

the method section of Pilot I. Participants who provided the incorrect direction for four or more of the six intervals, or who provided the corrected quality for all six intervals, had their data removed from main analyses.

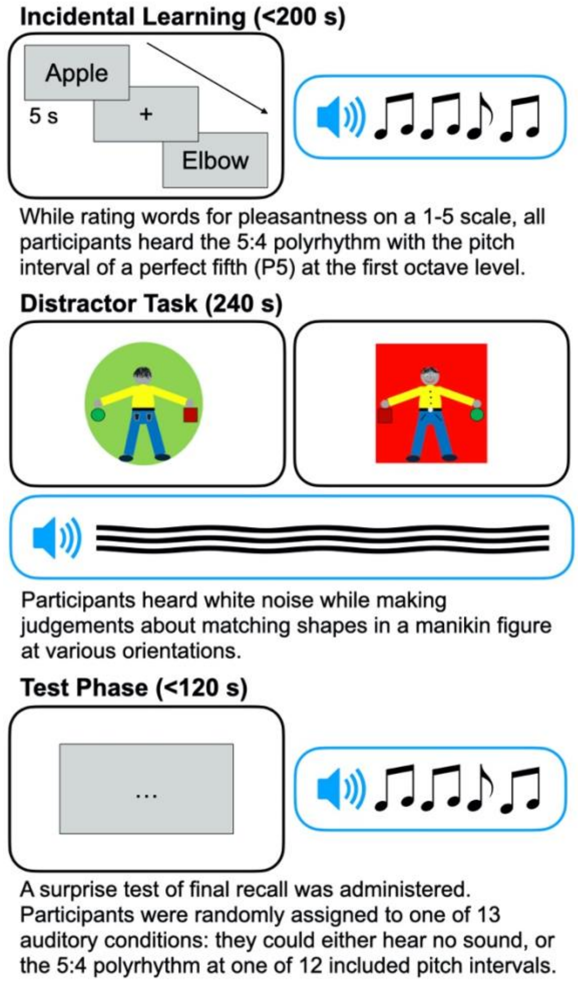
List Learning Task. Participants completed an adapted version of the List Learning Task (LLT) available in the online Millisecond Test Library (K. Borchert, 2017) in the Inquisit Web 6 player. The LLT consisted of a learning phase, a break, and a final test phase, over the course of which participants learned and were tested on 20 nouns selected from (Spreen & Schulz, 1966) norms, a random subset of those used by Mead and Ball (2007). These words were highly concrete, and varied in their emotional valence. See Appendix B for the full selection of words. For an overview of the LLT as adapted for present use, see Figure 6.

During the learning phase, words were presented individually for 5 seconds, followed by a 1 second fixation cross between each word. Participants were instructed to rate the pleasantness of a given word by pressing a number 1-5 on their keyboard, where 5 indicated a highly pleasant word, and 1 indicated a highly unpleasant word. Words were presented in a random order, with each word appearing twice.

Following a break, participants completed a surprise test of final free recall, during which participants were given two minutes or until they manually proceeded to recall as many words as possible, in any order they wished, by typing them into an on-screen text box.

Distractor Task. In order to provide an engaging break from learning words and limit rehearsal of material by participants, the Manikin Test of Spatial Orientation and Transformation, available in the online Millisecond Test Library, was used (K. Borchert, 2014). In this task, participants are shown a humanoid figure in one of several orientations (facing towards or away from participants, right-side-up or up-side-down), holding a small green circle in one hand, a small red square in the other, and positioned inside a larger version of one of these shapes. Participants are asked to evaluate in which hand the figure is holding a shape that matches the larger, surrounding shape. During a practice block, participants received feedback on their responses (in the center of the screen, “incorrect” appeared in red if they were incorrect; otherwise, “correct” appeared in green), followed by a fixation cross for 1 second before the next image. During the test block, no feedback was given, and the block lasted for 240 seconds.

Figure 6.
Procedure for the adapted list learning task.



Procedure

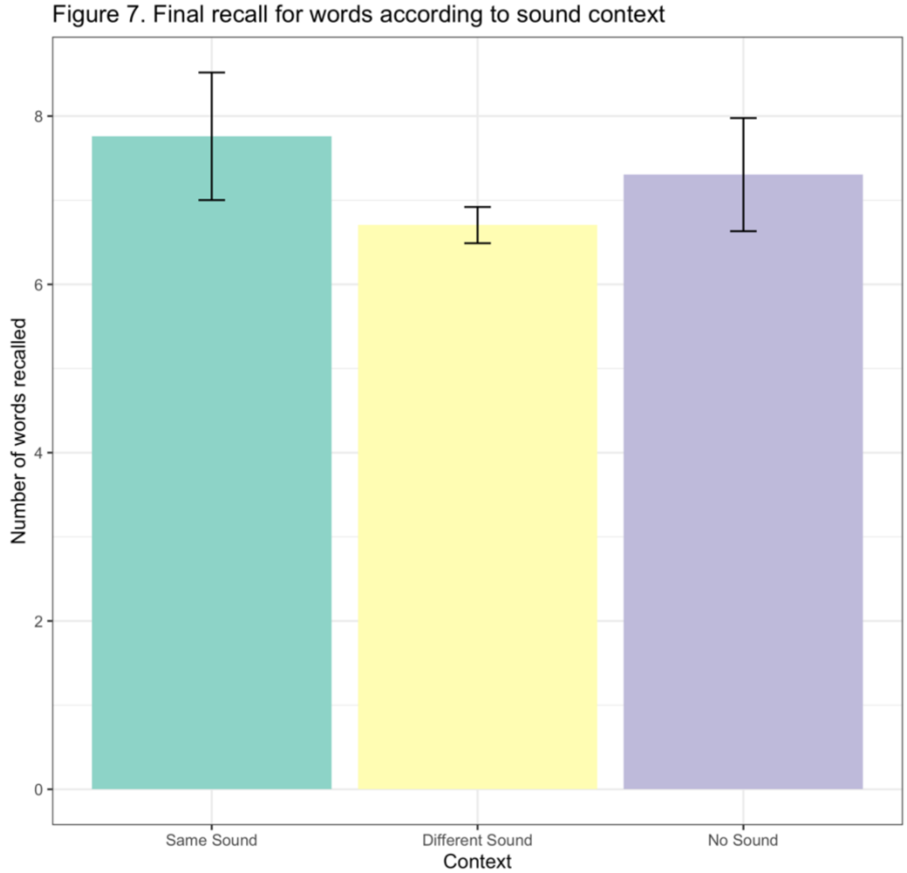
Participants who were determined to qualify and provided informed consent for the present study completed the experiment in the Inquisit Web 6 player. They first were prompted to listen to pink noise and set the volume at a comfortable level, which they were asked not to change throughout the task. Then, participants rated words for pleasantness while listening to the 5:4 polyrhythm at the interval of the perfect fifth (at the first octave level). After completing the ratings, they completed the practice and test blocks of the Manikin Test while listening to pink noise. Then, participants were given instructions for a surprise delayed test of recall on the words they had rated for pleasantness. During this final test, participants were randomly assigned to one of 13 conditions, determining what sound they heard for the test phase. Participants either heard no sound, or heard the musical stimuli at one of the 12 possible pitch intervals. Afterwards, participants completed the interval recognition test, reported whether they believed they heard the same sound while rating words and when tested on words, and submitted information about demographics and musical experience. Finally, they were thanked, debriefed, and provided instructions to receive payment.

Results

Musical Context

A one-way ANOVA tested the effect of the test sound context (either the same, a different, or no sound) on memory. This found no significant difference in the number of words recalled on average between members of different groups, $F(2,282) = 1.356$, $p = 0.259$. These results are displayed in Figure 7. A 3x4 factorial ANOVA tested whether final memory varied according to the octave level (first, second, or third) or interval class (major second, perfect fifth, major sixth, or perfect eighth) for those who heard a sound during the test period. Neither the main effect of octave level ($F(2,250) = 0.239$, $p = 0.788$) nor interval class ($F(3,250) = 0.416$, $p = 0.742$) reached significance, and their interaction was only marginally significant, $F(6,250) = 2.078$, $p = 0.056$. Exploratory post-

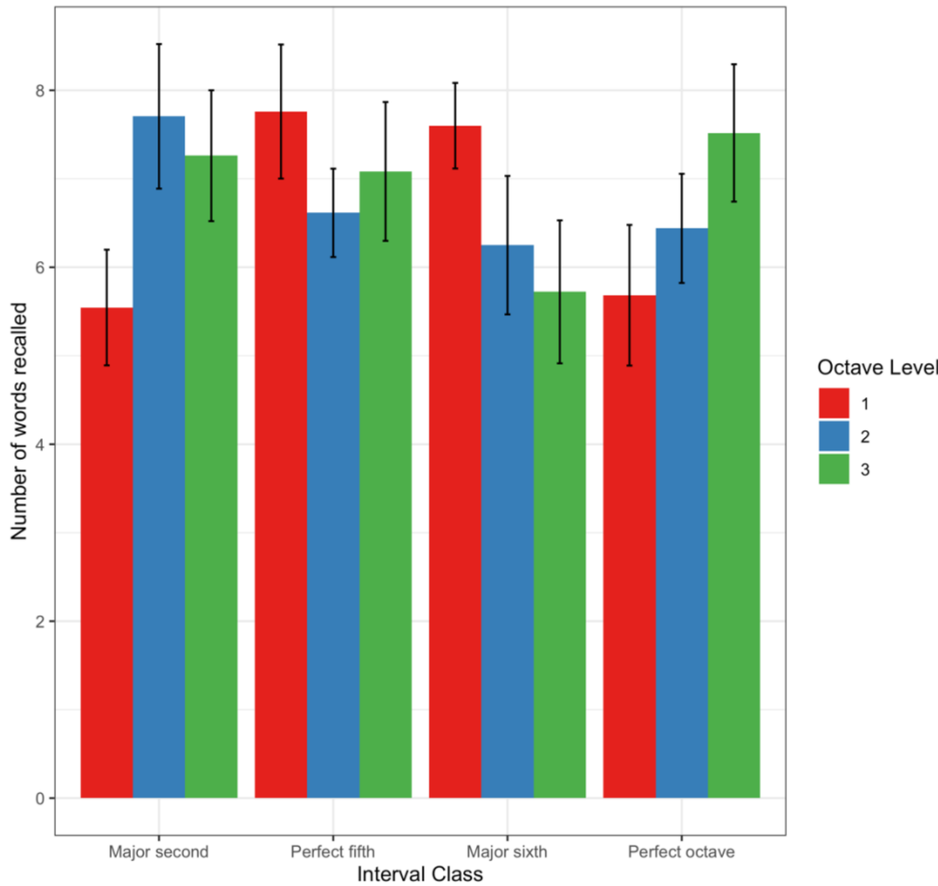
Figure 7. Final recall for words according to sound context



Mean words recalled by each group displayed. Bars indicate standard error of the mean (SEM).

hoc analyses of only the interaction term using a Tukey HSD test found no significant difference between pairs when correcting for multiple comparisons, all p_{adj}'s > 0.05. These results are displayed in Figure 8.

Figure 8. Final recall for words according pitch interval



Mean words recalled by each group displayed. Bars indicate standard error of the mean (SEM).

Reported Context

Pre-registered mediation analyses sought to investigate whether participants' explicit report of whether they heard the same sound during learning and test phases mediated the relationship between test sound and differences in memory performance. However, the results of the two ANOVAs indicate that there was no relationship between test sound and memory to be mediated. Instead of performing mediation analysis, I compared the effect of reported context (whether they thought the test sound was the same, different, or they were unsure compared to the earlier sound) and actual context (whether at test the same, a different, or no sound played) on memory performance using a 3x3 factorial ANOVA. There was a significant effect of reported context ($F(2,276) = 7.776$, $p = 5.19 \times 10^{-4}$), though neither the effect of actual context ($F(2,276) = 2.248$, $p = 0.107$) nor their interaction ($F(4,276) = 2.289$, $p = 0.06$) reached significance. A Chi-Square test of independence found that the number of participants across the nine possible combinations of actual and reported context did not vary significantly from expected values ($X^2(4, N = 285) = 6.903$, $p = 0.141$), and Table 8 shows descriptive values.

Table 8. Summary of participants by actual and reported condition.

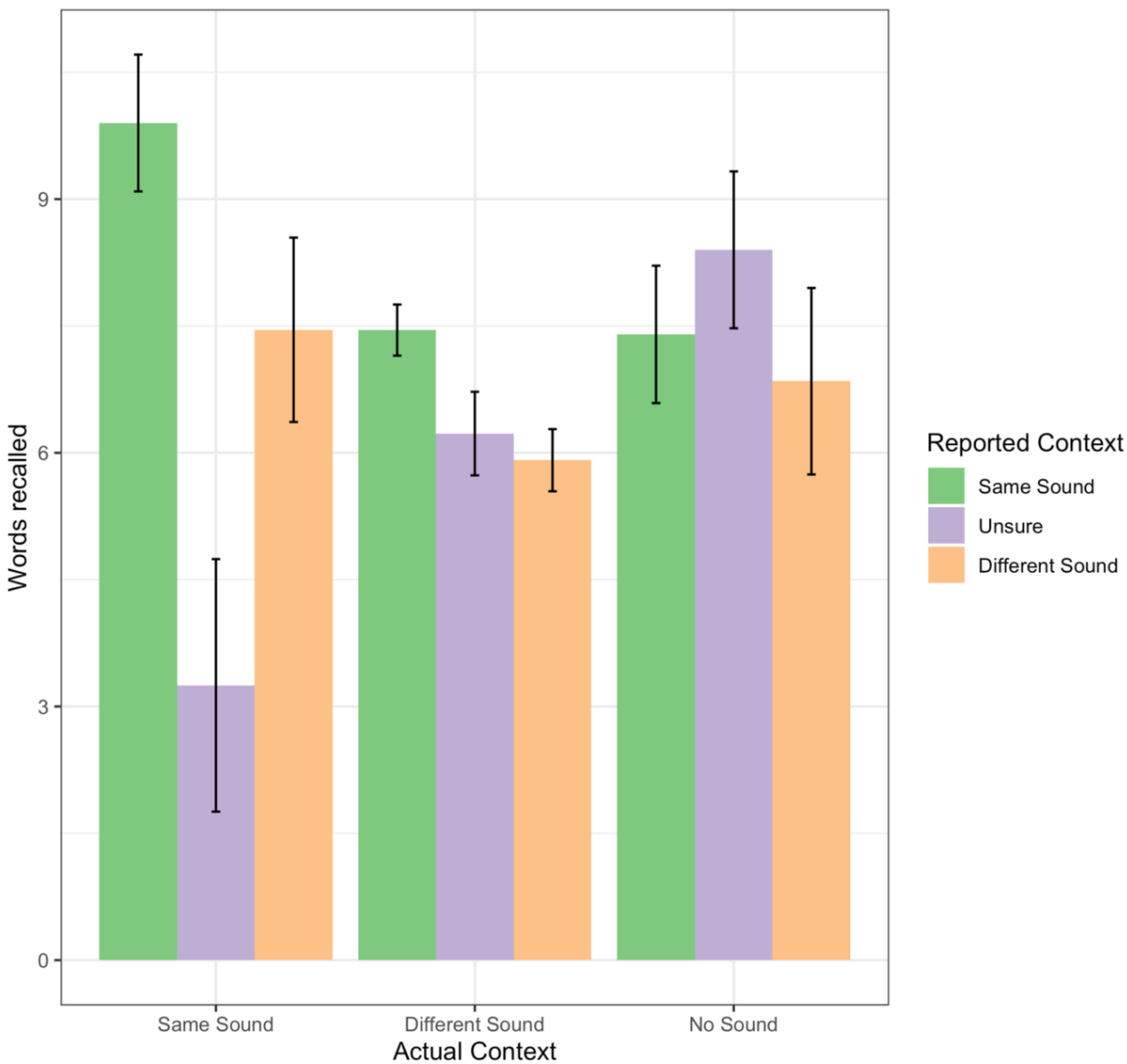
	n	Proportion
Same sound		
Reported same sound	10	40.00%
Reported unsure	4	16.00%
Reported different sound	11	44.00%
Different sound		
Reported same sound	113	47.68%
Reported unsure	44	18.57%
Reported different sound	80	33.76%
No sound		
Reported same sound	5	21.74%
Reported unsure	5	21.74%
Reported different sound	13	56.52%

Post-hoc analyses conducted with a Tukey HSD found that people who reported hearing the same sound recalled more words on average ($M = 7.641$, $SD = 3.177$) than those who reported hearing a different sound ($M = 6.192$, $SD = 3.419$, $p_{adj} = 2.38\text{x}10^{-3}$), or who were unsure ($M = 6.208$, $SD = 3.295$, $p_{adj} = 0.02$), controlling for what sound they actually heard during the test period. Additionally, four pairwise comparisons of the interaction between reported and actual context reached significance after controlling for multiple comparisons. Those who actually heard the same sound at test and who (correctly) reported hearing the same sound recalled significantly more words ($M = 9.9$, $SD = 2.558$) than participants belonging to any of three groups: those who actually heard the same sound but reported hearing a different sound ($M = 3.25$, $SD = 2.986$, $p_{adj} = 0.018$), those who actually heard a different sound and reported being unsure ($M = 6.227$, $SD = 3.277$, $p_{adj} = 0.037$), and those who actually heard a different sound and reported hearing a different sound ($M = 5.913$, $SD = 3.289$, $p_{adj} = 8.96\text{x}10^{-3}$). Participants who actually heard a different sound during the test but who (incorrectly) reported hearing the same sound also recalled more words ($M = 7.451$, $SD = 3.21$) than those who actually heard a different sound and reported hearing a different sound (M = 5.913, SD = 3.289, p_{adj} = 0.035). See Figure 9 for visualization of these results.

Moderation Analyses

Three simultaneous regressions were conducted to investigate the potential moderating effect of musical training or features of audio playback on the relationship between actual and reported context and memory performance. All models are summarized in Table 9. The first replicated the previous factorial ANOVA, evaluating the effect of actual and reported context on memory performance.

Figure 9. Final recall according to actual and reported musical context



Mean words recalled by each group displayed. Bars indicate standard error of the mean (SEM).

The second model added musical training as a factor in addition to those included in the first model. Participants reporting previous musical training did not explain any variance in the number of words recalled, *b* = 0.305, *p* = 0.437, and there were no significant interactions between musical training and either actual or reported context. The third model added two factors in addition to those included in the first model: whether participants reported altering the volume of audio playback at any point during the task (*b* = 0.185, *p* = 0.77), and whether participants reported listening to task audio through headphones or speakers (*b* = 0.377, *p* = 0.347), with neither significantly predicting differences in recall.

Table 9.

Summary of three linear regression models predicting number of words recalled.

		Model 1			Model 2			Model 3		
Variable	Level	Beta (*b*)	95% CI	*p*	Beta (*b*)	95% CI	*p*	Beta (*b*)	95% CI	*p*
(Intercept)		8.69	(7.32, 10.1)	<0.001**	8.57	(7.17, 9.98)	<0.001**	8.38	(6.63, 10.1)	<0.001**
Actual Context										
	Different	-1.18	(-2.54, 0.18)	0.089	-1.22	(-2.58, 0.14)	0.079	-1.23	(-2.6, 0.14)	0.078
	None	-0.174	(-2.04, 1.69)	0.855	-0.238	(-2.11, 1.64)	0.803	-0.311	(-2.21, 1.59)	0.747
Reported Context										
	Unsure	-1.49	(-2.54, -0.43)	0.006**	-1.45	(-2.51, 0.39)	0.007**	-1.49	(-2.55, -0.43)	0.006**
	Different	-1.57	(-2.43, -0.71)	<0.001**	-1.57	(-2.43, -0.71)	<0.001**	-1.61	(-2.48, -0.74)	<0.001**
Music Training										
	Yes				0.305	(-0.47, 1.08)	0.437			
Changed Volume										
	No							0.185	(-1.06, 1.43)	0.77
Audio Equipment										
	Speakers							0.377	(-0.41, 1.16)	0.347
		$R^2_{adj} = 0.47$			$R^2_{adj} = 0.046$			$R^2_{adj} = 0.043$		
		$F(8,283) = 4.527$			$F(9,282) = 3.7381$			$F(10,281) = 3.1618$		
		$p = 0.00147$			$p = 0.00272$			$p = 0.00513$		

Exploratory analyses

Sample Characteristics. Through descriptive analyses, the overall number of words recalled by participants in this experiment (*M* = 34.225%, *SD* =16.775%) was found to be lower than in comparable studies of music dependent memory, which tend to report average overall recall rates of 50% of learned words. In order to investigate sample characteristics that may have contributed to this difference, exploratory regression analyses tested whether age was a significant predictor of the number of words recalled. In a similar manner to the moderation analyses, a model with actual and reported context as predictors of recall was compared to one that added age as a predictor, with these models summarized in Table 10.

Table 10.

Predicting recall according to sound and participant features.

		Model 1			Model 4		
Variable	Level	Beta (*b*)	95% CI	*p*	Beta (*b*)	95% CI	*p*
(Intercept)		8.69	(7.32, 10.1)	<0.001**	10.03	(9.13, 10.93)	<0.001**
Actual Context							
	Different	-1.18	(-2.54, 0.18)	0.089	-1.112	(-1.8, -0.43)	0.106
	None	-0.174	(-2.04, 1.69)	0.855	-0.236	(-1.18, 0.71)	0.803
Reported Context							
	Unsure	-1.49	(-2.54, -0.43)	0.006**	-1.294	(-1.83, -0.76)	0.017*
	Different	-1.57	(-2.43, -0.71)	<0.001**	-1.363	(-1.81, -0.92)	0.002**
Age					-0.037	(-0.05, -0.02)	0.019*
		$R^2_{adj} = 0.47$					
		$F(8,283) = 4.527$					
		$p = 0.00147$					

Even controlling for the effects of actual and reported context on recall, age significantly predicted memory performance ($b = -0.037$, $p = 0.019$), such that younger participants could be expected to recall more words than older participants at a statistically significant but numerically small rate. In order to reliably predict at least one word to be additionally recalled by a younger participant, that participant would need to be at least 27 years younger than a participant otherwise matched in terms of actual and reported context.

Power. While the imbalance between group sizes in the preregistered comparison of actual test context's effect on memory performance was expected, the exclusion of a large number of participants based on their poor performance on the interval recognition task may have been problematic for the power in investigations of octave level and interval class, and the interaction between actual and reported context. There were as few as 16 participants in some pitch interval contexts (major sixth at the second octave level, $n = 16$; major second at the second octave level, $n = 17$; major sixth at the third octave level, $n = 18$), and as few as 4 participants in some combinations of actual and reported context (see Table 8). So, preregistered analyses were repeated, including participants who failed to correctly identify the direction of musical intervals more than half of the time during the interval recognition task.

In these analyses, neither the direction nor significance of any results differed from those conducted on the sample with preregistered exclusions. There was no difference in the number of words recalled according to whether participants heard the same, a different, or no sound ($F(2,342) = 1.39$, $p = 0.25$). For those who heard a sound at test, neither the octave level ($F(2,306) = 0.201$, $p = 0.818$), interval class ($F(3,306) = 0.393$, $p = 0.758$), nor their interaction ($F(6,306) = 1.322$, $p = 0.247$) had an effect on the number of words recalled. Finally, while memory performance was significantly different according to reported context ($F(2,336) = 10.9$, $p = 2.59\text{x}10^{-5}$) even when controlling for actual context, memory performance varied neither according actual context ($F(2,336) = 2.307$, $p = 0.101$) nor the interaction between actual and reported ($F(4,336) =1.573$, $p = 0.181$). Post-hoc analyses conducted with a Tukey HSD test of pairwise comparisons found that those who reported hearing the same sound at test as during learning recalled significantly more words ($M = 7.53$, $SD = 3.14$) than those who reported hearing a different sound ($M = 5.98$, $SD = 3.45$, $p_{adj} = 0.000383$) or being unsure ($M = 5.83$, $SD = 3.46$, $p_{adj} = 0.001250$). Additionally, those who correctly reported hearing a different sound at test recalled fewer words ($M = 5.694$, $SD = 3.36$) than those who either correctly reported hearing the same sound ($M = 9.0$, $SD = 3.303$, $p_{adj} = 0.030900$) or reported hearing the same sound but actually heard a different sound ($M = 7.4$, $SD = 3.144$, $p_{adj} = 0.003620$).

Of note, compared to the distribution of participants reported in Table 6, including poor performers in the interval recognition task did not include any new participants who heard the same sound but reported being unsure ($n = 4$) or hearing a different sound ($n = 11$), or any who heard no sound but reported hearing the same sound ($n = 5$), and this only included two new participants who heard no sound and reported being unsure ($n = 7$). The other 58 participants included in these analyses whose goal was to improve the tests' power were distributed among groups who already had larger numbers of participants represented.

Discussion

While there was no effect of either the presence or features of test sounds on memory performance, there was a significant effect of reported context on memory performance. Participants who reported hearing the same sound at test as they had when first exposed to the target words recalled significantly more words than other participants, even when controlling for whether participants actually heard the same, a different, or no sound. While participant age was a significant predictor of the number of words recalled, other participant characteristics such as musical experience or their method of audio delivery were not found to moderate the effects of actual or reported auditory context on memory performance.

A Remark on Methodology

To the best of the author's knowledge, this is the first study on music-dependent memory to be conducted online, and the first to include participants other than undergraduate students. Compared to these previous studies, this experiment:

- Had no control over the physical location or other contextual factors experienced by participants;
- Had no control over the method by which participants listened to the study's audio;
- Included older participants;
- Presented words twice during the incidental learning phase, which only some previous studies have done;
- Included a distractor task between the learning and test periods;
- Had that distractor task last for longer than in other studies;
- Did not vary the learning context between participants;
- Explicitly stated that the study involved sound;
- Used a novel range of musical stimuli.

Many of these differences were related to conducting the experiment online: namely, the lack of control over the environment compared to when in a physical lab, the difference in participant characteristics, the use of the distractor task as an attention check, and the necessity to facilitate a comfortable listening experience and informed consent by stating the use of sounds throughout the study. The use of a novel range of musical stimuli was an intentional difference, key to the study's goal of investigating the effect of pitch interval on musical perception. Relatedly, in order to maximize the observation of changes to this musical stimuli, participants were randomly assigned to hear different possible sounds (or no sound) during the test period, but participants did not hear different sounds during the learning context, intrinsically limiting the musical claims that could have been made by this study.

While previously cited studies tended to see average recall scores of about 50% of the learned words across all conditions, participants in the present study recalled an average of 6.845 words, or about 34% of the learned words. This high-level difference may be attributable to participant characteristics and the length of the distractor task. Exploratory analyses found that participants of greater age recalled fewer words at a numerically small but statistically significant rate. Other studies of music-dependent memory frequently included distractor sounds, but rarely included distractor tasks between the learning and test phases. Only Isarida et al. (2008) included a distractor task, which consisted of simple calculations for participants who had intentionally learned words, lasting for the length of time it took researchers to read test instructions to participants who had incidentally learned words. The increased cognitive effort involved in the current study's visuospatial distractor task may have increased the difficulty of recalling words during the subsequent surprise test (Barrouillet et al., 2007; Camos & Portrat, 2015). Additionally, the distractor task lasted for four minutes, longer than most previous studies' distractor periods (in one comparable study, participants listened to birdsound for 240 seconds, Mead & Ball, 2007), which may have accounted for small increases in participants' lessened ability to recall incidentally learned material.

Music-Dependent Memory

There was no effect of musical context on the number of words recalled during a surprise final test of recall. Participants who heard the same sound during this test did not recall more words than those who heard a different or no sound; additionally, those who at test heard sounds with an interval class of a perfect fifth did not recall significantly more words than other participants, and those whose sound was at the first octave level did not recall more words than other participants. This unexpected null effect may be explainable by both methodological and theoretical factors.

Methodologically, the variance in physical location and other contextual factors, study audio delivery method, number of word presentations, and explicit statement of sounds' role in the study could have affected the music-dependent memory for target information. Since participants were not all in the same location while completing the experimental task, and there is no guarantee that any given participant stayed in the same physical location throughout the task, it is difficult to account for the potential confounding effects of different ambient background contexts. Even within the same physical location, background disruptions or changes in the dynamic surrounding context could easily diminish the global effect of the study's musical context manipulation on a participants' perception of overall context (T. Isarida & Isarida, 2014; S. M. Smith, 1995; S. M. Smith & Vela, 2001). Participants were free to listen to the study's sounds however was most convenient for them, though the moderation analyses showed that neither changing the volume of playback nor listening to audio on speakers rather than headphones predicted differences in total recalled words.

Additionally, while Mead and Ball (2007) found significant effects of musical context for words presented twice to participants during the incidental learning phase, Isarida et al. (2008) performed a within-subject manipulation of word presentation, finding effects of musical context on the recall of words presented once, but not those presented twice. Isarida et al. (2008) hypothesized that twice-presented words have stronger representations in memory, but diminished associative connections to surrounding contextual features, diminishing the effects of musical context manipulation.

Past work in music-dependent memory frequently told participants that the background music they would hear during the task was present in order to make them more comfortable. This cover story may have diminished the extent to which participants paid attention to the sounds, and along with the ecological validity of the musical excerpts used by other researchers may have contributed to a perception of the music as background music. The online nature of the present study, and the fact that the included sounds a) were not rated as remarkably pleasant by participants in the pilot experiments, and b) do not possess structural similarity to typical background music, complicated the presentation of a similar cover story. Participants who qualified after completing a screener task on MTurk titled “Answer questions about sound and language” could then 1-10 days later complete the experimental task titled “Listen to sounds while assessing words and pictures.” Full descriptions of the tasks as seen by participants are included in the IRB proposals included in Appendix D. The differences in initial description of the sounds, in conjunction with features of the sounds themselves, could have led to the sounds in this study being considered as target information in a similar manner to the words presented during the study.

Transitioning to more theoretical explanations, both global- and feature-level evaluations of musical context may have been made difficult by the time delay between the learning and test timepoints, the fine-grained manipulation in the present study, and the social acclimation to features of the sounds. Accuracy in recognition judgements between the relative size of intervals diminish as the time between two target intervals increases (E. M. O. Borchert, 2011; McPherson & McDermott, 2020; Prince et al., 2009). If we assume that the same perceptual bottleneck limits these explicit recognition judgements as would at least in part limit any nonconscious evaluation of contexts used during a retrieval process, the length of the distractor task may have made it more difficult for participants to compare the learning and test contexts during the final test of recall. Both inaccurate judgements and an ambiguous representation of musical context could have diminished the role of context in affecting memory.

Even if the learning and test sounds were recognizably different, the present study's manipulation may not have been sufficient to facilitate evaluations of the different sound as a new context. The global impression of the context could have been influenced by the characteristics of the pitch interval. However, the pilot study supported the selection of the major second, perfect fifth, major sixth, and perfect octave because these intervals were rated as similar to each other on extramusical features such as familiarity, pleasure, and distraction. So, these differences may have been diminished to the extent that the measured extramusical features were critical to that global impression. Consistent with past work in mental context theory, even sounds with a different holistic representation due to differences in pitch interval, feature-level similarities may still have allowed for contextual benefits from different sounds. The sounds in this study were exactly the same in terms of overall tempo, MIDI soundfont, playback volume, use of the 5:4 polyrhythm, its four-note component being played by the lower note in the interval, and this low note being the same pitch and register, a C4 (262 Hz). It is therefore easy to imagine that of the abundance of features common between two sounds at a different pitch interval, some number of these features may have been critical to context reinstatement during test; this feature-level similarity between different sounds may have weakened the effect of this single feature's difference.

Finally, polyrhythms and five-limit tuning are musically interesting and rooted in notions of resonance and low-integer ratio representations of harmonic series relationships that likely influence our perception of most features of the sounds around us (Chew, 2001; Large & Snyder, 2009; van Noorden & Moelants, 1999). They are not, however, standard features in American popular music. Past work has demonstrated interactions between pitch and rhythm on musical perception, and work by Moelants and van Noorden (2005) suggested that the 5:4 polyrhythm at 150 bpm provided a set of metric constrains at which the different pitched components were of equal salience to each other. However, their data was gathered from a musically trained population, likely to have more exposure or at least tolerance to novel sounds than the average American resident. While reporting musical training didn't moderate the effect of actual and reported context on memory performance, nor did it predict whether participants would report being in the same or a different context, there may have been additional differences between the characteristics or experiences of this sample compared to those in previous music cognition studies. It is difficult to measure the extent to which these sounds were more abnormal than the musical pieces selected in previous studies of music-dependent memory, but they likely were perceived as more abnormal. In particular, the looped polyrhythm may have, given its infrequency in American popular music, have been the most unfamiliar and therefore most salient feature of the musical stimuli, further limiting the extent to which the global impression of these sounds could be affected by altering the pitch interval.

No Sound Is as Good as Any

While these rationale may explain the lack of difference in words recalled by people who heard different sounds, it fails to explain why those who heard no sound during their test did not recall significantly fewer words than other participants, as was expected. One reassuring remark is that participants who heard no sound during their test did not recall significantly more words than those who heard sound, which is consistent with past research in music-dependent memory suggesting that the presence of sound during the test is not reliably distracting or disrupting recall processes.

This study also relied on only a single possible learning sound, unlike other context-dependent memory studies which randomly assign participants to their learning and test contexts. While this was a useful decision to include a high number of participants exposed to each test sound, it limits the ability to make any general claims about the change in context experienced by participants. That is, the no sound condition in this study did not vary the learning sound experienced by participants, so these data only inform us as to how hearing no sound at test *after hearing a perfect fifth at learning* affects recall. It remains possible that participants who heard a major second while learning, for example, could have demonstrated a greater difference in words recalled between people who heard the major second again versus no sound during their test. Future work with similar stimuli would benefit from randomly assigning participants to both a learning and test context in order to make broader claims, in order to make claims about hearing no sound at test robust to what particular sound was heard during learning.

A related drawback of this study is the lack of control over or insight into the retrieval strategies used by participants during their final test. Context information even latently related to previously learned information can be an effective retrieval tool (Karpicke et al., 2014; Long et al., 2015; Whiffen & Karpicke, 2017). Some participants may have facilitated the reinstatement of previous context, including thinking of previously heard sounds, in order to aid their recall of target words. This particular strategy could have been easier for those who heard no sound at test compared to those who heard different sounds. Additionally, the retrieval of any words may have strengthened the representation of the sound heard during learning. Any combination of these possible occurrences could have minimized differences in final recall between participants in the same- and no-sound conditions. Randomly assigning both learning and context conditions would make it possible to compare memory of those who heard no sound at both learning and test, compared to those who heard some sound during learning.

Actual Versus Reported Context

One unexpected outcome of this study was the effect of reported context on the number of words recalled during the final test. Participants who reported hearing the same sound during both learning and test phases recalled significantly more words than those who reported the sound being different or those who reported being unsure. Reported context did not depend on actual context, as a chi square test indicated that the frequency at which people reported these contexts did not differ according to what sound they actually heard.

It is possible that more is captured by reported context than a true reflection of participants' evaluations of the learning and test sounds' similarity. One argument in favor of this response reflecting noise is the high rates of error in participants' reports. Fewer than half of participants who did hear a sound at test correctly identified it as either the same or a different sound. Over a fifth of participants who heard no sound during the test reported hearing the same sound as when they were rating the sounds for pleasantness. Participants who heard no sound but who were asked to reflect on the test sound – “Think back to the sound you heard while rating words for pleasantness, and the sound you heard when you were later tested on those words. Did you hear the same sound both times?” – may rightfully have been confused. The question could have been designed to explicitly recognize that at test, participants may have heard the same, a different, or no sound at all. While participants may also have not paid close attention to the question, originally analyzed data only included participants who passed the Manikin and interval recognition task attention checks, diminishing the likelihood of responses being consistently inattentive.

Additionally, though this question was asked immediately after participants finished the surprise test of recall, the act of reflecting on both the learning and test period may have encouraged participants to update their representation of the sounds heard at both timepoints. Participants may not have consciously compared the sounds until asked to do so here, and their reports could have been influenced by motivations to have considered the sounds as similar or dissimilar, according to ideas about the study’s goals or the likelihood that they were supposed to hear different sounds. The use of retrieval strategies that relied on contextual cues may have updated their representation of the musical context at test in a way that conscious reflection could allow for the test sound to be considered similar to the earlier learning sound, even if there was no sound during a participant’s test period. These issues of conscious reflection and the boundaries of accurate recognition are similar to those inherent in previous studies in psychoacoustics and music cognition, as discussed in the introduction.

Reported context may therefore not capture an infallible evaluation of two sounds' similarity, but it is reasonable to assume that it may accurately reflect participants' belief, upon reflection, about their similarity. With this framework, these results suggest that music-dependent memory effects were contingent primarily on participants' belief that the musical context was the same during a test of target information as when they were originally exposed to the information. Imagined contexts have been found to produce context-dependent memory effects to at least the same extent as "real" contexts (Masicampo & Sahakyan, 2014; S. M. Smith & Vela, 2001). This could be explained through two mechanisms. First, a *belief* that you are in the same context, characterized by a conviction in one's perceptual assessment of two circumstances informing a holistic context representation that is the same in both instances. Second, even if someone doesn't *believe* they are in the same context, reinstatement of key features of a context can still strengthen the accessibility of related target information, and could strengthen the connective representation between the context and learned material.

Further conclusions about the interaction between actual and reported context are more complicated to interpret. Robust to the inclusion of participants who performed poorly on the interval recognition task, participants who correctly believed they heard the same sound at both timepoints outperformed several other groups on the final test of recall. However, in both the preregistered and exploratory set of participants, there were very few participants representing certain combinations of actual and reported context. Since groups' recall scores were consistently found to have equal variance according to Bartlett tests of homogeneity of variance, the ANOVAs were likely still robust despite the imbalance of group sizes (Grace-Martin, 2020). However, the power of these analyses were constrained by the size of the smallest included groups; since as few as 4 or 5 participants were in several included groups, this study was inadequately powered to report on the complex interaction between these factors.

Future Directions

The goal of the present study was to investigate the extent to which the context-dependent memory paradigm could be a useful method for indirectly assessing the perceptual similarities between different sounds. It is possible that the interaction of methodological factors such as the distraction period's task and length, displaying words twice during the incidental learning phase, and the narrow ways in which the musical stimuli were altered contributed to the null results of this study. Continued efforts to assess perceptual similarity through this indirect method may still be rich, given the complexity of reported context's effect on memory performance, and the interactions between actual and reported context that were detectable in this sample. In order to refine the methodology used for the indirect assessment of perceptual similarity, future research may benefit from shorter or less intensive distractor tasks, and more systematically evaluating whether the number of word presentations has an effect on the strength of context manipulations.

Additionally, future work in this vein should balance the tradeoffs between varying both learning and test contexts for participants and investigating the myriad ways in which the sounds themselves could be altered. Varying both learning and tests contexts would allow researchers to draw claims about how altering features of a musical context affects judgements of sounds' similarity, *regardless of the original sound*. However, there are more features of musical significance, and certainly more of perceptual significance, than the pitch of the high and fast note in 5:4 polyrhythm. Investigations into pitch interval would benefit from varying the actual pitches used in order to make claims about pitch interval more broadly, rather than different intervals constructed above C4; at the very least, altering whether the two notes of an interval are played in the 4- or 5-note component of the polyrhythm would allow for generalization beyond the case where the C4 is always the lower, 4-note component of the sound. Given future knowledge of effective methodology to pursue this question, tempo or polyrhythmic density manipulations would be musically rich and extend previous work in music-dependent memory (Balch et al., 1992; Balch & Lewis, 1996; T. K. Isarida et al., 2017).

The unexpected effect of reported context suggests equally rich lines of future inquiry. Of course, future work that varies both learning and test context may still struggle to ensure distributions of participants across actual and reported context categories to allow for more effectively powered analyses. However, this would allow for claims to be made about how reported context is or is not influenced by manipulations of sound, rather than changes from a particular original sound. It is possible that work with a less cognitively intense and/or shorter delay period between learning and test might find a different relationship between actual and reported context. As participants are more able to make accurate recognition judgements between sounds, reported context may mediate this effect, and rates of error in reported context may be lower. Alternatively, if beliefs about the perceptual similarity of sounds are informed by more than the physical features of a sound – even when those physical features are more easily recognized – reported context may still independently effect final recall scores. Similarly, we may find that participants who incorrectly report hearing a different sound outperform those who correctly report hearing a different sound; that is, while belief in a similar sound may reflect context effects being present during learning, the same physical context may still facilitate context-based memory effects even absent the belief that the contexts were similar.

References

Ames, C. (1989). The Markov Process as a Compositional Model: A Survey and Tutorial. *Leonardo*, *22*(2), 175–187.

Angelis, V., Holland, S., Upton, P. J., & Clayton, M. (2013). Testing a Computational Model of Rhythm Perception Using Polyrhythmic Stimuli. *Journal of New Music Research*, *42*(1), 47–60. https://doi.org/10.1080/09298215.2012.718791

Balch, W. R., Bowman, K., & Mohler, L. A. (1992). Music-dependent memory in immediate and delayed word recall. *Memory & Cognition*, *20*(1), 21–28. https://doi.org/10.3758/BF03208250

Balch, W. R., & Lewis, B. S. (1996). Music-Dependent Memory: The Roles of Tempo Change and Mood Mediation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(6), 1354–1363.

Ball, L. J., Shoker, J., & Miles, J. N. V. (2010). Odour-based context reinstatement effects with indirect measures of memory: The curious case of rosemary. *British Journal of Psychology*, *101*(4), 655–678. https://doi.org/10.1348/000712609X479663

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, *39*(3), 445–459. https://doi.org/10.3758/BF03193014

Barrouillet, P., Bernardin, S., Portrat, S., Vergauwe, E., & Camos, V. (2007). Time and cognitive load in working memory. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *33*(3), 570–585. https://doi.org/10.1037/0278-7393.33.3.570

Ben-Haim, M. S., Eitan, Z., & Chajut, E. (2014). Pitch memory and exposure effects. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(1), 24–32. https://doi.org/10.1037/a0033583

Blistein, J. (2020, March 18). Katy Perry Wins Appeal in ‘Dark Horse’ Infringement Case. *Rolling Stone*. https://www.rollingstone.com/music/music-news/katy-perry-dark-horse-copyright-win-appeal-969009/

Borchert, E. M. O. (2011). *Effects of Timing and Context on Pitch Comparisons between Spectrally Segregated Tones* [PhD Dissertation]. University of Minnesota.

Borchert, K. (2014). *User Manual for Inquisit’s Manikin Test of Spatial Orientation and Transformation*. Millisecond Software, LLC. https://www.millisecond.com/download/library/v6/mentalrotation/manikintest/manikintest/manikintest.manual

Borchert, K. (2017). *User Manual for Inquisit’s List Learning Task*. Millisecond Software, LLC. https://www.millisecond.com/download/library/listlearningtask/

Borchert, K. (2020). *User Manual for Inquisit’s Lexical Decision Task*. Millisecond Software, LLC. https://www.millisecond.com/download/library/v6/lexicaldecisiontask/lexicaldecisiontask/lexicaldecisiontask.manual

Camos, V., & Portrat, S. (2015). The impact of cognitive load on delayed recall. *Psychonomic Bulletin & Review*, *22*(4), 1029–1034. https://doi.org/10.3758/s13423-014-0772-5

Cann, A., & Ross, D. A. (1989). Olfactory Stimuli as Context Cues in Human Memory. *The American Journal of Psychology*, *102*(1), 91–102. https://doi.org/10.2307/1423118

Cases from 2010-2019. (n.d.). Music Copyright Infringement Resource. Retrieved April 22, 2021, from https://blogs.law.gwu.edu/mcir/cases-2/2010-2019/

Chew, E. (2001). *Modeling Tonality: Applications to Music Cognition*.

Cousineau, M., Carcagno, S., Demany, L., & Pressnitzer, D. (2014). What is a melody? On the relationship between pitch and brightness of timbre. *Frontiers in Systems Neuroscience*, *7*. https://doi.org/10.3389/fnsys.2013.00127

Cousineau, M., Demany, L., & Pressnitzer, D. (2009). What makes a melody: The perceptual singularity of pitch sequences. *J. Acoust. Soc. Am.*, *126*(6), 9.

Demany, L., Pressnitzer, D., & Semal, C. (2009). Tuning properties of the auditory frequency-shift detectors. *The Journal of the Acoustical Society of America*, *126*(3), 1342–1348. https://doi.org/10.1121/1.3179675

Demany, L., & Ramos, C. (2005). On the binding of successive sounds: Perceiving shifts in nonperceived pitches. *J. Acoust. Soc. Am.*, *117*(2), 9.

Demany, L., & Semal, C. (2007). *The Role of Memory in Auditory Perception* (Vol. 29, pp. 77–113). https://doi.org/10.1007/978-0-387-71305-2_4

Demany, L., & Semal, C. (2018). Automatic Frequency-Shift Detection in the Auditory System: A Review of Psychophysical Findings. *Neuroscience*, *389*, 30–40. https://doi.org/10.1016/j.neuroscience.2017.08.045

Dowling, W. J. (1978). Scale and contour: Two components of a theory of memory for melodies - ProQuest. *Psychological Review*, *85*(4), 341–354.

Dowling, W. J., & Harwood, D. L. (1986). *Music Cognition*. Academic Press.

Echaide, C., Río, D. del, & Pacios, J. (2019). The differential effect of background music on memory for verbal and visuospatial information. *The Journal of General Psychology*, *146*(4), 443–458. https://doi.org/10.1080/00221309.2019.1602023

Eich, E. (1980). The cue-dependent nature of state-dependent retrieval. *Memory & Cognition*, *8*(2), 157–173.

Eich, E. (1995). Mood as a mediator of place dependent memory. *Journal of Experimental Psychology: General*, *124*(3), 293–308. https://doi.org/10.1037/0096-3445.124.3.293

Eich, E., & Metcalfe, J. (1989). *Mood Dependent Memory for Internal Versus External Events*. 13.

Ellis, M. C. (1991). An Analysis of “Swing” Subdivision and Asynchronization in Three Jazz Saxophonists. *Perceptual and Motor Skills*, *73*(3), 707–713. https://doi.org/10.2466/pms.1991.73.3.707

Frolova-Walker, M. (1998). “National in Form, Socialist in Content”: Musical Nation-Building in the Soviet Republics. *Journal of the American Musicological Society*, *51*(2), 331–371. https://doi.org/10.2307/831980

Gann, K. (2019). *The arithmetic of listening: Tuning theory and history for the impractical musician*. University of Illinois Press.

Godden, D. R., & Baddeley, A. D. (1975). Context-Dependent Memory in Two Natural Environments: On Land and Underwater. *British Journal of Psychology*, *66*(3), 325–331. https://doi.org/10.1111/j.2044-8295.1975.tb01468.x

Goldman, A., Jackson, T., & Sajda, P. (2020). Improvisation experience predicts how musicians categorize musical structures. *Psychology of Music*, *48*(1), 18–34. https://doi.org/10.1177/0305735618779444

Grace-Martin, K. (2020, December 18). When Unequal Sample Sizes Are and Are NOT a Problem in ANOVA. *The Analysis Factor*. https://www.theanalysisfactor.com/when-unequal-sample-sizes-are-and-are-not-a-problem-in-anova/

Graves, J. E., & Oxenham, A. J. (2017). Familiar Tonal Context Improves Accuracy of Pitch Interval Perception. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.01753

Howard, M. W., & Kahana, M. J. (2002). A Distributed Representation of Temporal Context. *Journal of Mathematical Psychology*, *46*(3), 269–299. https://doi.org/10.1006/jmps.2001.1388

Inquisit 6 Web. (2020). https://www.millisecond.com

Isarida, T., & Isarida, T. K. (2014). *Environmental context-dependent memory* (pp. 115–152).

Isarida T. K., Isarida T., & Hayashibe K. (2008). *Context dependent effects of background music on the free recall of incidentally and intentionally learned words*. https://doi.org/10.5265/JCOGPSY.5.107

Isarida, T. K., Kubota, T., Nakajima, S., & Isarida, T. (2017). Reexamination of Mood-Mediation Hypothesis of Background-Music-Dependent Effects in Free Recall. *Quarterly Journal of Experimental Psychology*, *70*(3), 533–543. https://doi.org/10.1080/17470218.2016.1138975

Jafari, M., & Ansari-Pour, N. (2019). Why, When and How to Adjust Your P Values? *Cell Journal (Yakhteh)*, *20*(4), 604–607. https://doi.org/10.22074/cellj.2019.5992

Jones, A. (2001). Mass Music and the Politics of Phonographic Realism. In *Yellow Music: Media Culture and Colonial Modernity in the Chinese Jazz Age* (pp. 105–136).

Justus, T., & Bharucha, J. (2002). *Music Perception and Cognition* (pp. 453–492). https://doi.org/10.1002/0471214426.pas0111

Karpicke, J. D., Lehman, M., & Aue, W. R. (2014). Chapter Seven - Retrieval-Based Learning: An Episodic Context Account. In B. H. Ross (Ed.), *Psychology of Learning and Motivation* (Vol. 61, pp. 237–284). Academic Press. https://doi.org/10.1016/B978-0-12-800283-4.00007-1

Kimmel, W. (1941). Vaughan Williams's Melodic Style. *The Musical Quarterly*, *27*(4), 491–499.

Kleinsmith, A. L., & Neill, W. T. (2018). Recognition of transposed melodies: Effects of pitch distance and harmonic distance. *Psychonomic Bulletin & Review*, *25*(5), 1855–1860. https://doi.org/10.3758/s13423-017-1406-5

Large, E. W., & Kolen, J. F. (1994). Resonance and the Perception of Musical Meter. *Connection Science*, *6*(1), 177–208.

Large, E. W., & Snyder, J. S. (2009). Pulse and Meter as Neural Resonance. *Annals of the New York Academy of Sciences*, 12.

Lavengood, M. (2019). What Makes It Sound '80s?The Yamaha DX7 Electric Piano Sound. *Journal of Popular Music Studies*, *31*(3), 73–94. https://doi.org/10.1525/jpms.2019.313009

Lavengood, M. L. (2007). *A New Approach to the Analysis of Timbre*. City University of New York (CUNY).

Lehman, S. H. (2012). *Liminality as a Framework for Composition: Rhythmic Thresholds, Spectral Harmonies and Afrological Improvisation* [Columbia University]. https://doi.org/10.7916/D8RJ4RKM

Levin, T. (2013). Making Marxist-Leninist Music in Uzbekistan. In *Music and Marx: Ideas, Practices, Politics* (pp. 190–2003).

Litman, L., Robinson, J., & Abberbock, T. (2017). TurkPrime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods*, *49*(2), 433–442. https://doi.org/10.3758/s13428-016-0727-z

London, J. (2002). Cognitive Constraints on Metric Systems: Some Observations and Hypotheses. *Music Perception - MUSIC PERCEPT*, *19*, 529–550. https://doi.org/10.1525/mp.2002.19.4.529

Long, N. M., Danoff, M. S., & Kahana, M. J. (2015). Recall dynamics reveal the retrieval of emotional context. *Psychonomic Bulletin & Review*, *22*(5), 1328–1333. https://doi.org/10.3758/s13423-014-0791-2

Manning, J. R., Polyn, S. M., Baltuch, G. H., Litt, B., & Kahana, M. J. (2011). Oscillatory patterns in temporal lobe reveal context reinstatement during memory search. *Proceedings of the National Academy of Sciences*, *108*(31), 12893–12897. https://doi.org/10.1073/pnas.1015174108

Masicampo, E. J., & Sahakyan, L. (2014). Imagining another context during encoding offsets context-dependent forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(6), 1772–1777.

McDermott, J. H., Keebler, M. V., Micheyl, C., & Oxenham, A. J. (2010). Musical intervals and relative pitch: Frequency resolution, not interval resolution, is special. *The Journal of the Acoustical Society of America*, *128*(4), 1943–1951. https://doi.org/10.1121/1.3478785

McPherson, M., & McDermott, J. H. (2020). Time-dependent discrimination advantages for harmonic sounds suggest efficient coding for memory. *Proceedings of the National Academy of Sciences*, *117*(50), 32169–32180.

Mead, K., & Ball, L. (2007). Music tonality and context-dependent recall: The influence of key change and mood mediation. *European Journal of Cognitive Psychology - EUR J COGN PSYCHOL*, *19*, 59–79. https://doi.org/10.1080/09541440600591999

Moelants, D., & van Noorden, L. (2005). The Influence of Pitch Interval on the Perception of Polyrhythms. *Music Perception*, *22*(3), 425–440. https://doi.org/10.1525/mp.2005.22.3.425

Mu, Y. (1994). Academic Ignorance or Political Taboo? Some Issues in China's Study of Its Folk Song Culture. *Ethnomusicology*, *38*(2), 303–320. https://doi.org/10.2307/851742

Mullan, E. (n.d.). *Nine most notorious copyright cases in music history*. Retrieved April 22, 2021, from https://www.bbc.com/culture/article/20190605-nine-most-notorious-copyright-cases-in-music-history

Neto, P., Cui, A.-X., Rojas, P., Vanzella, P., & Cuddy, L. (2021). Not just cents: Physical and psychological influences on interval perception. *Psychomusicology: Music, Mind, and Brain.* http://dx.doi.org/10.1037/pmu0000272

Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review*, *116*(1), 129–156. https://doi.org/10.1037/a0014420

Prince, J. B., Schmuckler, M. A., & Thompson, W. F. (2009). The effect of task and pitch structure on pitch-time interactions in music. *Memory & Cognition*, *37*(3), 368–381. https://doi.org/10.3758/MC.37.3.368

Rothstein, R. A. (1980). The Quiet Rehabilitation of the Brick Factory: Early Soviet Popular Music and Its Critics. *Slavic Review*, *39*(3), 373–388. https://doi.org/10.2307/2497160

Sandbank, L. (2019). *To Conceive of Consonance in Chaos: The Influence of the Harmonic Series on the Perception of a New Musical System*. Bard College.

Schweer, W. (2020). *Musescore* (3.5.2) [Computer software]. https://musescore.org

Smith, D. M., Wakeman, D., Patel, J., & Gabriel, M. (2004). Fornix Lesions Impair Context-Related Cingulothalamic Neuronal Patterns and Concurrent Discrimination Learning in Rabbits (Oryctolagus cuniculus). *Behavioral Neuroscience*, *118*(6), 1225–1239. https://doi.org/10.1037/0735-7044.118.6.1225

Smith, S. M. (1985). Background Music and Context-Dependent Memory. *The American Journal of Psychology*, *98*(4), 591–603. JSTOR. https://doi.org/10.2307/1422512

Smith, S. M. (1995). Mood Is a Component of Mental Context: Comment on Eich (1995). *Journal of Experimental Psychology: General*, *124*(3), 309–310. https://psycnet.apa.org/doi/10.1037/0096-3445.124.3.309

Smith, S. M. (2009). The hippocampus, context processing, and episodic memory. In E. Dere, A. Easton, L. Nadel, & J. Huston (Eds.), *Handbook of Episodic Memory* (Vol. 18, pp. 465–481). Elsevier.

Smith, S. M., & Vela, E. (2001). Environmental context-dependent memory: A review and meta-analysis. *Psychonomic Bulletin & Review*, *8*(2), 203–220. https://doi.org/10.3758/BF03196157

Snyder, J. S., Carter, O. L., Lee, S., Hannon, E. E., Alain, C., Centre, B., & Care, G. (2008). *Effects of Context on Auditory Stream Segregation*.

Spreen, O., & Schulz, R. W. (1966). Parameters of abstraction, meaningfulness, and pronunciability for 329 nouns. *Journal of Verbal Learning and Verbal Behavior*, *5*(5), 459–468. https://doi.org/10.1016/S0022-5371(66)80061-0

Stark, S. M., Reagh, Z. M., Yassa, M. A., & Stark, C. E. L. (2018). What's in a context? Cautions, limitations, and potential paths forward. *Neuroscience Letters*, *680*, 77–87. https://doi.org/10.1016/j.neulet.2017.05.022

Suchoff, B. (1972). Bartók and Serbo-Croatian Folk Music. *The Musical Quarterly*, *58*(4), 557–571.

Tari, L. (2006). Bartók's Collection of Hungarian Instrumental Folk Music and Its System. *Studia Musicologica Academiae Scientiarum Hungaricae*, *47*(2), 141–166.

Tenney, J. (2015a). Hierarchical Temporal Gestalt Perception in Music: A Metric Space Model (with Larry Polansky)(1978). In L. Polansky, L. Pratt, R. Wannamaker, & M. Winter (Eds.), *From Scratch* (pp. 201–233). University of Illinois Press. https://www.jstor.org/stable/10.5406/j.ctt6wr6t7.14

Tenney, J. (2015b). META Meta ≁ Hodos (1975). In L. Polansky, L. Pratt, R. Wannamaker, & M. Winter (Eds.), *From Scratch* (pp. 166–179). University of Illinois Press. https://www.jstor.org/stable/10.5406/j.ctt6wr6t7.12

Tenney, J. (2015c). On the Physical Correlates of Timbre (1965). In L. Polansky, L. Pratt, R. Wannamaker, & M. Winter (Eds.), *From Scratch* (pp. 128–131). University of Illinois Press. https://www.jstor.org/stable/10.5406/j.ctt6wr6t7.9

Tse-tsung, M. (1956). *Chairman Mao's Talk to Music Workers*. Selected Works of Mao Tse-tsung Volume 7. https://www.marxists.org/reference/archive/mao/selected-works/volume-7/mswv7_469.htm

van Noorden, L., & Moelants, D. (1999). Resonance in the Perception of Musical Pulse. *Journal of New Music Research*, *28*(1), 43–66. https://doi.org/10.1076/jnmr.28.1.43.3122

Velasco, M. J., & Large, E. W. (2011). Pulse Detection in Syncopated Rhythms using Neural Oscillators. *Poster Session*, 6.

Wang, A. X., & Wang, A. X. (2020, January 9). How Music Copyright Lawsuits Are Scaring Away New Hits. *Rolling Stone*. https://www.rollingstone.com/pro/features/music-copyright-lawsuits-chilling-effect-935310/

Whiffen, J. W., & Karpicke, J. D. (2017). The role of episodic context in retrieval practice effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(7), 1036–1046. https://doi.org/10.1037/xlm0000379

Williams, R. V., Kidson, F., Fuller-Maitland, J. A., Sharp, C. J., & B, L. E. (1906). Songs Collected from Norfolk. *Journal of the Folk-Song Society*, *2*(8), 161–183.

Zarate, J. M., Ritson, C. R., & Poeppel, D. (2012). Pitch-interval discrimination and musical expertise: Is the semitone a perceptual boundary? *The Journal of the Acoustical Society of America*, *132*(2), 984–993. https://doi.org/10.1121/1.4733535

Appendix A. Musical Stimuli

Figure A1.

Figure A1. The 36 diatonic intervals between a minor second and three octaves are shown to the left, at three octave levels and 12 interval classes. The interval classes are named according to the interval name most commonly used in musical analysis (e.g., Major third) as well as the frequency ratio used to construct (and tune) the interval (e.g., 5:4).

Figure A2. The 12 pitch intervals used in the experiment, at the interval classes of major second, perfect fifth, major sixth, or perfect octave, and at the three octave levels.

Appendix B. Words

Table B1.

fouling	sadist	belch	cinders	relapse	gazer	conveyor	mulch	broach	eyesore	spatter	pacers
huntsman	clasping	shackles	warhorse	forceps	curbside	casanova	amoeba	huntress	preying	netting	tiebreak
thespian	halogens	hunches	grubs	gluttony	dropout	hoaxes	wobbling	laments	clank	typhoid	figurine
marigold	haystack	pansies	liqueur	grating	pennants	nellie	aerosols	tonsil	padlock	greenery	irritant
gobble	huddling	ravine	defector	shrines	curlers	parables	truckers	pewter	aviators	subtypes	salami
goblet	tipoff	minks	savanna	Cromwell	gymnast	archway	parishes	spiky	libido	blotch	pulpit
notches	mince	doubter	armpits	scaffold	mallard	hemming	prowl	nines	pennant	sameness	slinger
hoodlums	jurists	laxative	scrawl	tricolor	rawhide	errand	rigging	cellars	recluse	hernia	thrones
cavities	trumps	flops	macaroni	tumbler	tartar	gertrude	tantrums	forester	crevices	mongoose	palmetto
bulkhead	ironside	rambles	lioness	steed	tricycle	upshot	dominoes	beehive	costumer	kepler	poacher
riser	abrasion	footpath	longings	prancing	sinning	scorch	sardine	caddie	wrangle	clucking	jiffy
mayans	mower	domes	ratty	eminence	muskets	trances	peekaboo	amnesty	gobbler	absorber	phoney
duckling	tweezer	stances	rosary	felon	capes	rigidity	hiccup	stopover	belching	molars	heirloom
trachea	cholera	frontage	adjuncts	lingual	tidings	clemency	grocers	heckle	furnaces	gainer	firmer
knitwear	killjoy	sulfide	affix	enormity	stencil	lawmen	lullaby	liaisons	pizarro	slumming	cobble
coves	wholes	pranks	saunter	muzzles	eyelets	mobster	tomcat	psyches	medics	laymen	capers
ducky	mainstay	ladle	cutlery	monet	fussing	taunts	pastes	chairing	auntie	noontime	toppling
stanza	nudism	glycerol	Saxons	knickers	omelette	starling	chomp	eclair	igloo	outcrop	khaki
repast	caprice	parasol	peeler	pricks	sundial	spyglass	splicing	crucible	cobwebs	unreason	twine
burlap	sinuses	squatter	rogues	bumming	starlet	mamma	matting	drugging	lassie	pooch	pollock
debuts	hoosier	croquet	swarms	scooting	macbeth	gardenia	crutches	mishap	drawl	emergent	plazas
snobbery	manger	antacid	wedlock	radish	oracles	mussels	buffets	tolls	glycerin	entrails	serene
rivet	bonfires	agonies	bruiser	loaves	crumple	tinsel	curds	culprits	heaves	figment	reproach
veiling	Shawnee	swagger	clincher	gurgle	swish	tracings	middles	friar	crockery	havens	couplet
envoys	bristles	gleam	scuff	breather	campsite	skids	burnside	inning	tinfoil	screech	nuance
grope	windbag	hawkers	moped	mucus	antidote	dueling	sadism	ringside	recitals	pulley	sawing
vinimize	clooping	vertin	aghost	moftware	pellboy	gandhold	yellop	triend	pragnet	famisn	bricycle
lummar	estract	etuity	oplivion	abalyst	wimpid	greakout	nebruska	hobago	vartical	indeen	clith
bagans	deprime	closated	absinte	gragies	iller	bottem	repirted	heptile	sharf's	beavened	croadly
voicus	peresy	liftors	carthabe	toplin	johms	pamages	elmiba	tharaoh	stumblod	ampaling	oppusing
dalaces	raisley	palisape	amenue	joshuo	pissile	seasting	abbured	bedsare	scropes	traffed	henedict
idonclad	pummelad	nitrois	gustre	soners	pervoded	prinnied	rostpone	dineties	mencil	whytll	audobahn
lomby	hygry	unreado	trome	Leonurd	banor	agounded	lecrete	whola	hansies	helict	elitor
redonism	talmid	warblor	somedian	chonetic	crislo	acylum	crofound	prall	Dobertan	satrons	tassie
magoons	bussycat	trightly	adsords	hecades	gonesome	sathrobe	cheriffs	poonybin	imsult	mockney	delfare
sprart	righways	satchmen	lobstem	sceptle	peatnik	gnowy	potruck	cromote	pauntify	inforno	edduring
uddue	oufright	bollage	boignant	edocate	chetched	merdant	hibacha	brivado	hartin	campfure	culletin

laotiad	ladyshir	glamiel	emucator	aspiros	preaming	walkaver	newrite	henchant	veunite	pexation	arnesty
chatever	udilize	pumin	survoys	horridge	palving	pranners	warmip	ethanop	Jou's	uncarled	legree
bespots	tabrics	sminking	grandel	aphorast	drozzly	minerja	wastige	puraty	panine	crostate	treamt
barflime	minarity	belapse	cuttors	dopside	morldly	ilhibit	fastade	spifty	parklire	amenily	sorona
croothly	teadman	emecy	girearm	osidize	priant	atostle	hesuit	neafiest	agongst	bracture	editome
triving	welleg	teriodic	streens	caystack	florast	rotutes	walterp	otsidian	domanly	holition	romunia
polygin	fentally	beappear	greyve	clowfly	unmixud	linalist	hileage	tayrors	palivary	livilian	fondone
teagued	videshow	avateur	medicel	parpoon	vaselone	nourash	tridont	supertly	bourneys	miredly	slatd
bentler	hempter	leeing	erongate	edich	horribic	sidewose	imtegers	posaic	mutinoos	smapper	cration
sophasm	ecoligy	twipping	loolness	polony	tarrison	varios	launer	codiatry	asimal	imparse	crounds
carony	plickers	sebacle	matisfy	instunct	artenna	ownselp	epists	cadwoman	pandsome	tanities	synamses
exderly	gandler	fociable	jobdess	bedrick	soldrums	geinous	grecks	erthrone	mavender	fluggish	sidnoys
lonety	geadow	clyke	salmin	lutcher	sarvests	purmise	dylak	conniter	gyanide	emclave	waximum
aprans	ditied	bothic	turdened	shambers	flackest	gransly	poktains	ivnite	fickbed	veddlers	nockpits
collequy	sairless	reenager	antedape	fathens	puties	puckily	pleavage	decruits	snylight	anitator	envign

Table B1. List of all words and nonwords included in the lexical decision task of the pilot studies. A total of 312 words (top 26 rows) and 312 nonwords (bottom 26 rows) were randomly selected from the 367 nouns and 1147 nonwords that the English Lexicon Project provided as comparable in length and task performance in reaction time and accuracy.

Table B2.

knife	cousin	witness	gift
flame	professor	tribe	council
basket	prison	library	cheek
sheep	card	bush	nurse
maid	flour	apple	wool

Table B2. List of all words used in the experiment for ratings of pleasantness and later recall. Words were a random subset of those used by Mead and Ball (2007).

Appendix C. IRB Materials

Entry C1. IRB Proposal, initial submission, November 5, 2020.

Section 1

1. **Today's date:** Wednesday, November 4, 2020
2. **Name:** Hadley Parum
3. **Email:** hp4041@bard.edu
4. **Your Academic Program/Department/Office:** Psychology Department
5. **Your status (faculty, staff, graduate or undergraduate student):** Undergraduate student, Senior I
6. **Adviser or Faculty Sponsor (if applicable):** Justin Hulbert
7. **If you are a graduate or undergraduate student, has your Adviser or Faculty Sponsor seen and approved your application?** Yes
8. **Your Adviser's or Faculty Sponsor's email address (if applicable):** jhulbert@bard.edu
9. **Please list all individuals (full name and status, i.e. faculty, staff, student) involved in this project that will be working with human subjects. Note: Everyone listed must have completed Human Subject Research Training within the past three years.** Hadley Parum
10. **Do you have external funding for this research?** Maybe. (Submitted as "No" on the form.)
11. **If so, state the name of the sponsor and the title of the project as it was submitted to that sponsor.** I applied for external funding from the Psi Chi Honor Society on October 15, 2020. A response about a grant of up to $1,500 is expected to be received by December 15, 2020. The title as submitted was "Varying Pitch Interval as Cued Musical Context Demonstrates the Perceptual Similarity of Polyrhythms".

Section 2

1. **What is the title of your project?** Varying Pitch Interval as Cued Musical Context Demonstrates the Perceptual Similarity of Polyrhythms
2. **When do you plan to begin this project? (Start date):** December 1, 2020
3. **Describe your research question(s):** We know that pitch and rhythm interact to explain our perception of music. However, it is not fully understood what aspects of these key features are most essential to the perceptual similarity of musical examples. To investigate this further, the present research relies on a test of context-dependent memory of English words using a test of delayed free recall. Participants will learn words while hearing a polyrhythm played at the pitch interval of a perfect fifth, and will later be tested on the studied words while hearing the polyrhythm at the same or a varied pitch interval (e.g., the minor second, major third, and major sixth), or will hear no additional sound. It is expected that participants will recall a higher proportion of words correctly if the musical-auditory context is the same at test as it was during learning. Furthermore, I expect recall to be higher if participants heard a) pitch intervals in the same octave level as that heard at learning, such that intervals are closer rather than farther apart, and if they heard b) an interval class consonant to the original interval or the ratio of the polyrhythm. These findings aim to show that perceptual relationships like similarity depend on certain relational aspects of key features, like proximity and consonance, furthering understanding of perception and music cognition.
4. **Describe the population(s) you plan to recruit and how you plan to recruit participants. Please submit all recruitment material, emails and scripts to IRB@bard.edu.** Participants will be recruited from the Bard community and through Amazon's Mechanical Turk (MTurk). Before they agree to participate, they will be determined to be above the age of 18, with English

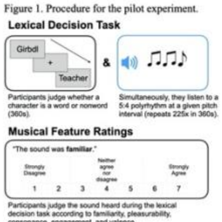
as a dominant language, and without self-reported hearing impairments or perfect pitch; MTurk participants will also be restricted to those currently residing in the United States. Additionally, all participants will verify that the device used for the experiment has functional audio output and the capacity to download the Inquisit Web 6 player,[1] which is compatible with most Windows, Mac, iOS, and Android devices. All participants will be informed that the experiment will involve listening to audio and downloading the Inquisit program before required to begin the experiment. Any participants recruited through MTurk will be provided a description of the task's purpose, length, and use of audio playback as well as the Inquisit Web player, and will be able to volunteer through Amazon's website to participate in exchange for disclosed monetary compensation. Any participants recruited from the Bard community may be solicited through publicity measures such as posters in campus locations or through social media.

5. **Will your participants include individuals from vulnerable or protected populations (e.g., children, pregnant women, prisoners, or the cognitively impaired)?** No.
6. **If your participants will include individuals from the above populations, please specify the population(s) and describe any special precautions you will use to recruit and consent.** N/A.
7. **Approximately how many individuals do you expect to participate in your study?** Total estimate: 804 unique participants. Pilot study: 24 participants; screener questionnaire: 780 participants; main experiment: 390 participants.
8. **Describe the procedures you will be using to conduct your research. Include descriptions of what tasks your participants will be asked to do, and about how much time will be expected of each individual. NOTE: If you have supporting materials (printed surveys, questionnaires, interview questions, etc.), email these documents separately as attachments to IRB@bard.edu. Name your attachments with your last name and a brief description (e.g., "WatsonSurvey.doc).**
   a. Pilot Study
      i. Summary: A pilot study will be conducted to measure how potential musical stimuli vary on extra-musical characteristics, in order to better inform the stimuli used during the main experiment. Figure 1 provides a visual overview of the procedures discussed here. Participants recruited from the Bard community will be asked to complete a 45-minute task entirely on the Inquisit Web 6 platform on an independent device. The time window during which the participant takes the study will be scheduled with the primary investigator, but the scheduling and completion of the experiment will not require the participant to physically meet the primary investigator at any point. The participant will complete the screening questions and informed consent measure (see Appendices D and B), followed by a short interval recognition task, described below. Then, participants will repeat a short listening procedure ten times. This listening procedure consists of completing the lexical decision task, described below while hearing a given musical stimuli on loop, and then in silence completing the musical feature ratings for the given audio (see Appendix D), where they answer Likert-type questions expressing dis/agreement with statements such as "The sound was familiar." Following that approximately 30 minute listening procedure, they will complete short

[1] View download page and compatibility statement: https://www.millisecond.com/download/inquisitweb6.aspx

demographic and musical experience questionnaires (see Appendix D). Finally, they will be thanked, debriefed, and will receive payment.

Figure 1. Procedure for the pilot experiment.

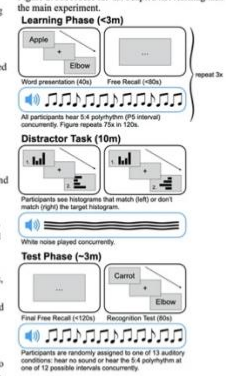


      ii. *Interval Recognition Task.* In order to confirm participant self-reports about hearing loss, perfect pitch and pitch deafness, and working audio equipment, they will complete a short (less than one minute) interval recognition task. Six musical examples, consisting of two notes played in sequence at 150bpm (for a five second musical example), will be presented to participants. After each one, they will be asked whether the second note was higher or lower than the first. Then, they will be asked to guess which interval name best fits the interval they heard between the two notes (e.g., minor second, perfect fourth, or major sixth).
      iii. *Musical Stimuli.* Created with Musescore, sounds will consist of 5:4 polyrhythms played at a given pitch interval, ranging from a unison (where both lines are played on the same note) to three octaves apart, for 37 total possible intervals included in the pilot experiment. Based on pilot experiment results, 12 of these intervals will be included in the main experiment. See Appendix F for examples of musical stimuli in music notation and for links to audio examples.
      iv. *Lexical Decision Task.* Participants will complete a task adapted from that which is publicly available on Inquisit's millisecond database,[2] in order to provide a within-subject measure of attention while evaluating words. They will be asked to categorize a given set of letter characters as either comprising an English word or a nonword, responding with an "I" key for valid words and an "E" for nonwords. Each word is presented briefly, for a period of 250ms, followed by a fixation cross in the center of the screen until their response, which is followed by 700ms of buffer time in which only the fixation cross is present. Including short test trials to acclimate participants to the controls, the presentation of all words takes approximately three minutes.
   b. Screener Questionnaire
      i. Summary: A screener questionnaire is a procedure supported by the community of researchers and MTurk workers in order to equitably determine eligible participants for future research, without disqualifying MTurk workers who would otherwise attempt to complete a study for which they are ineligible. Participants recruited through MTurk

[2] View the online list of publicly available tasks: https://www.millisecond.com/download/library/lexicaldecisiontask/

who complete the informed consent agreement will then complete the screening questionnaire and the demographics questionnaire (see Appendix D) in the Inquisit Web 6 platform. Following this, they will be thanked, debriefed, and directed to return to MTurk with a code to ensure payment. The whole procedure will take approximately two minutes to complete.

Figure 2. Procedure for the adapted list learning task in the main experiment.



   c. Main Experiment
      i. Summary: Participants on MTurk who completed the Screening task and are determined to qualify will be solicited directly through MTurk's contact system and invited to participate in an approximately 25-minute task. Figure 2 provides a visual overview of the procedures discussed here. Those who join and complete the informed consent agreement will then complete the interval recognition task (described above, under the pilot experiment). Following this, all participants will complete the learning portion of the List Learning Task, described below, while listening to the same musical stimuli (the 5:4 polyrhythm at a predetermined consonant interval, e.g. the perfect fifth). Participants will then listen to white noise for the duration of a ten minute visual distractor task. Afterwards, participants will complete the final tests of the List Learning Task while randomly assigned to one of 13 conditions, in which they may hear no audio, or one of the 12 musical stimuli included in the main experiment. Finally, participants will complete demographics and musical experience questionnaires (see Appendix D), and then will be thanked, debriefed, and directed to return to MTurk with a code to ensure payment.
      ii. *List Learning Task.* In order to measure delayed free recall and recognition of learned material, the Inquisit List Learning Task[3] will be adapted for present use in Inquisit 6 Web (2020). Participants will learn and be tested on the same 20 unrelated, neutrally valenced English words for objects (e.g., "elbow," "apple"). The task consists of a learning phase, in which the words are presented in fixed order every two seconds, followed by a short free recall period without feedback, three times. After a ten minute

[3] View the List Learning Task's code, user manual, and demo: https://www.millisecond.com/download/library/listlearningtask/

break or distractor task, there is a test of delayed recall lasting two minutes, where participants are asked to recall as many words as they can, in any order. Finally, to test delayed recognition, 20 words will be presented and participants will use key responses to report whether the word was in the studied list ("Y") or not ("N"). Half of the final 20 words were on the original list of 20 words, and the other half are 10 unstudied words.

iii. *Distractor Task.* Participants will complete a 10-minute visuospatial reasoning task, in order to provide a break between learning and test that occupies attentional resources, both to engage participants and limit rehearsal. To this end, participants will complete a task such as the Manikin Test[4] of Spatial awareness, in which participants must make judgements about which hand of a figure holds the same shape as that which surrounds the figure, or the Spatial Processing Task[5], in which participants must judge whether a histogram is a transformed version of a previously presented target histogram, or a differently shaped histogram.

9. **Describe any risks and/or benefits your research may have for your participants.**
   a. Risks: All components of this study involve minimal risk for participants. Due to the online nature of all tasks, there exists a small risk for discomfort or fatigue during stimulus delivery. For those involved in the pilot study and main experiment, the listening tasks may pose the additional small risk of auditory discomfort. To alleviate these risks, participants will a) be told in advance about the occurrence and length of exposure to these stimuli, b) be able to set the volume to a comfortable level before listening tasks, and c) be offered short, self-directed breaks at various points of the procedures. Since listening to audio and using electronic devices is a frequent occurrence for most participants, it is likely that they will be familiar with their comfort level during such tasks and be able to make informed decisions about participation and breaks.
   b. Benefits: This study offers no direct benefits to participants. They may receive indirect emotional benefits as a result of the knowledge that they have contributed towards scientific inquiry and the completion of a Senior Project. All participants will receive compensation for their time, according to applicable minimum hourly wages at the time of their participation:
      i. Participants recruited from the Bard community for the pilot study who complete the experiment before December 31, 2020 will receive $8.85 (for a 45-minute task and a New York State minimum wage of $11.80/hour).
      ii. Participants recruited from the Bard community for the pilot study who complete the experiment after December 31, 2020 will receive $9.38 (for a 45-minute task and a New York State minimum wage of $12.50/hour).
      iii. Participants recruited from MTurk for the screening questionnaire will receive $0.24 (for a 2-minute task at the U.S. Federal minimum wage of $7.25/hour).
      iv. Participants recruited from MTurk for the main experiment will receive $3.02 (for a 25-minute task at the U.S. Federal minimum wage of $7.25/hour).

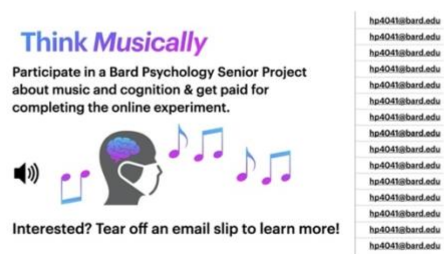
[4] View Manikin Test of Spatial Orientation and Transformations code, user manual, and demo: https://www.millisecond.com/download/library/mentalrotation/manikintest/

[5] View the Spatial Processing Task's code, user manual, and demo: https://www.millisecond.com/download/library/mentalrotation/spatialprocessingtask/

10. **Describe how you plan to mitigate (if possible) any risks the participants may encounter.** Since listening to audio and using electronic devices is a frequent occurrence for most participants, it is likely that they will be familiar with their comfort level during such tasks. Since listening to audio and using electronic devices is a frequent occurrence for most participants, it is likely that they will be familiar with their comfort level during such tasks. In order to further defray any potential discomfort, participants will:
    a. Be told in advance about the occurrence and length of exposure to these visual and auditory stimuli;
    b. Be able to set the volume to a comfortable level during the screening procedure, and
    c. Be offered short, self-directed breaks at various points of the procedures.
11. **Describe the consent process (i.e., how you will explain the consent form and the consent process to your participants):** All participants will be presented with the relevant consent form prior to the delivery of the task through Inquisit. The form contains general information about the goals of the experiment, risks and benefits, compensation, and confidentiality details, as well as the contact information for the primary investigator and the Bard Institutional Review Board.
12. **Have you prepared a consent form(s) and emailed it as an attachment to IRB@bard.edu?** Yes.
13. **If you are collecting data via media capture (video, audio, photos), have you included a section requesting consent for this procedure(s) in your consent form(s)?** N/A.
14. **If your project will require you to employ a verbal consent process (no written consent forms), please describe why this process is necessary and how verbal consent will be obtained and stored.** N/A.
15. **What procedures will you use to ensure that the information your participants provide will remain confidential and safeguarded against improper access or dissemination?** Participants will be assured during the consent and debriefing processes that steps are being taken to ensure the safeguarding of any personal information. Information collected through the Inquisit Web platform (answers to questionnaires and data from experimental tasks) will not be connected to their identifying information (for in-person participants, the name on their consent form; for MTurk participants, any information associated with their MTurk account) after the conclusion of data collection. All information provided through the Inquisit Web platform will be temporarily stored on their servers in Oregon, USA. Their security statement is accessible here. While it is stored there, it is only accessible to the primary investigator through their password-protected Inquisit account. At the conclusion of data collection, data will be stored offline on a password-protected computer to which only the primary investigator has access. Potentially personal information collected (e.g., responses to the screening questionnaire) will only be used to determine qualification for further participation, and not connected to their scores in the main experiment, and all data (including demographics and other questionnaires, and task performance) will be reported in aggregate.
16. **Will it be necessary to use deception with your participants at any time during this research? Withholding details about the specifics of one's hypothesis does not constitute deception, this is called incomplete disclosure. Deception involves purposefully misleading participants about the nature of the research question or about the nature of the task they will be completing.** No.

17. **If your project study includes deception, please describe here the process you will use, why the deception is necessary, and a full description of your debriefing procedures.** N/A.
18. **For all projects, please include your debriefing statement. (This is information you provide to the participant at the end of your study to explain your research question more fully than you may have been able to do at the beginning of the study.) All studies must include a debriefing statement. Be sure to give participants the opportunity to ask any additional questions they may have about the study.** See Appendix C. for the debriefing statements.
19. **If you will be conducting interviews in a language other than English, will you conduct all of the interviews yourself, or will you have the assistance of a translator? If you will be using the assistance of a translator, that individual must also certify that he or she is familiar with the human subject protocol and has completed the online training course.** N/A.
20. **If your recruitment materials or consent forms will be presented in languages other than English, please translate these documents and email copies to IRB@bard.edu.** N/A.

APPENDIX A: RECRUITMENT MATERIALS
Sample On-Campus Recruitment Materials



Sample Amazon Mechanical Turk (MTurk) Human Intelligence Task (HIT) Descriptions

Screening Questionnaire
Title: Auditory Experiences
Description: This study contains short multiple-choice questions about your experiences and attitudes about hearing, music, and language. The study will run in Inquisit Web 6. The activity serves as a screener for future research affiliated with the Bard College Psychology department to be conducted by the Primary Investigator.
Custom keywords: n/a
Time Estimate: 2 minutes.
Payment: $0.24USD

Experimental Task
Title: Sound and Cognition Study
Description: This study asks you to listen to various simple audio recordings while learning and manipulating verbal and nonverbal data presented visually, and asks short multiple-choice questions about your experiences and attitudes. This study will run in Inquisit Web 6.
Custom keywords: n/a
Time Estimate: 25 minutes.
Payment: $3.02USD

APPENDIX B: CONSENT FORMS
Pilot Experiment Consent Form
INFORMED CONSENT AGREEMENT

You are being asked to participate in a study designed as a part of a Bard College Senior Project in the Department of Psychology. This experiment seeks to investigate how sound and learning interact.

Please take time to thoroughly read through this form, which describes potential risks and benefits of participation in this study. After you have been fully informed, you have the right to choose whether you wish to participate by signing or not signing this form. You also have the ability to end your participation in this study at any time during or after your completion of the study.

Background: This experiment seeks to investigate how sound and learning interact. In this pilot study, you will be asked to listen to various sounds while making judgements about visually presented items as well as features of the sounds. This task is estimated to take approximately 45 minutes in total.

Risks and Benefits: Participating in this study poses minimal risk. As a part of this experiment, you would be asked to complete a series of tasks on the Inquisit Web 6 platform, which is free and safe to download, and compatible with most desktop and mobile devices. You will be asked to listen to different sounds, and will be allowed to determine a comfortable listening volume to avoid potential discomfort. You will also be asked to pay attention to items presented visually through the screen, and will be offered the opportunity to take short breaks periodically to alleviate the potential fatigue or discomfort resulting from time spent in front of a digital device.

While participation in this study may not provide any direct benefits to you, you will receive compensation for your time, described below. Additionally, you will provide crucial information that enhances our understanding of learning and human cognition, and you will support the completion of a Senior Project, therefore supporting an undergraduate student in developing research skills.

Compensation: Your compensation has been set in accordance with NY State minimum wage at your time of participation. [IF PARTICIPATING BEFORE December 31, 2020] Therefore, you will receive $8.85 for your time. [IF PARTICIPATING AFTER December 31, 2020] Therefore, you will receive $9.38 for your time.

Confidentiality: Information you provided through the Inquisit platform, including responses to questionnaires and your performance on tasks, will be temporarily stored on Inquisit's servers, and at the conclusion of data collection will be stored offline on password-protect computers to which only the primary investigator and their advisor will have access. These data will be connected to an anonymous subject number and kept separately from this consent form, which contains your name.

The final published version of this research will be permanently and publicly available as a Senior Project at the Stevenson Library of Bard College, and digitally through the Digital Commons. This information may also be used in the potential publication or presentation on findings that result from this project. In

these cases, your data will be presented in aggregate with the data of other participants and will not be linked to any identifiable information.

Your Rights as a Participant: Your participation in this experiment is entirely voluntary, meaning that you can choose to not participate or to withdraw from the experiment at any time without penalty. If you choose to withdraw, you will still have the opportunity to receive compensation for the time you planned to spend involved in the study. If you do wish to withdraw at any time, feel free to inform your experimenter, who can pause the experiment and withdraw your data from collection and analysis.

The experimenter will tell you more about the experiment at the completion of this session. If you have any further questions regarding this study, you may ask them during the session, or you may email the principal investigator, Hadley Parum (hp4041@bard.edu), the advisor overseeing this project, Justin Hulbert (jhulbert@bard.edu), or the Bard College Institutional Review Board (irb@bard.edu).

STATEMENT OF CONSENT:
"The purpose of this study, procedures to be followed, and the risks and benefits have been explained to me. I have been given the opportunity to ask questions, and my questions have been answered to my satisfaction. I have been told whom to contact if I have additional questions. I have read this consent form and agree to be in this study, with the understanding that I may withdraw at any time."

By checking the box below and proceeding, you are indicating the following:

- I am at least 18 years of age.
- I have fully read and understand the contents of this informed consent agreement.
- I agree with the above statement of consent.

[ ] I agree with the above statements and provide my informed consent to participate in this study.

Screener Questionnaire Consent Form (MTurk)
INFORMED CONSENT AGREEMENT

You are being asked to participate in a survey designed as a part of a Bard College Senior Project in the Department of Psychology. This survey seeks to learn about your experiences with hearing, music, and language, and will be used to screen participants for future research opportunities.

Please take time to thoroughly read through this form, which describes potential risks and benefits of participation in this study. After you have been fully informed, you have the right to choose whether you wish to participate by signing or not signing this form. You also have the ability to end your participation in this study at any time during or after your completion of the study.

Background: This survey seeks to learn about your experiences with hearing, music, and language. You will be presented with six multiple choice questions in the Inquisit Web 6 platform. The results will be used to screen individuals, determining their eligibility for different future research projects. This task is estimated to take two minutes to complete.

Risks and Benefits: Participating in this study poses minimal risk. As a part of this experiment, you would be asked to complete a series of tasks on the Inquisit Web 6 platform, which is free and safe to download, and compatible with most desktop and mobile devices. While you will be asked to attend to questions presented visually on-screen, the short length of the task is intended to alleviate the potential fatigue or discomfort resulting from time spent in front of a digital device.

While participation in this study may not provide any direct benefits to you, you will receive compensation for your time, described below. Additionally, you will provide crucial information that enhances our understanding of learning and human cognition, and you will support the completion of a Senior Project, therefore supporting an undergraduate student in developing research skills.

Compensation: For your participation, you will be provided with direct monetary compensation through MTurk totalling $0.24 for a 2-minute task.

Confidentiality: Information you provided through the Inquisit platform, including responses to questionnaires and your performance on tasks, will be temporarily stored on Inquisit's servers, and at the conclusion of data collection will be stored offline on password-protect computers to which only the primary investigator and their advisor will have access. These data will be connected to an anonymous subject number unrelated to your MTurk account. The completion code you receive at the completion of the task will be randomly issued by Inquisit and used only to disburse payment, and at no point will Inquisit receive information connected to your MTurk account.

The final published version of this research will be permanently and publicly available as a Senior Project at the Stevenson Library of Bard College, and digitally through the Digital Commons. This information may also be used in the potential publication or presentation on findings that result from this project. In these cases, your data will be presented in aggregate with the data of other participants and will not be linked to any identifiable information.

Your Rights as a Participant: Your participation in this experiment is entirely voluntary, meaning that you can choose to not participate or to withdraw from the experiment at any time without penalty.

The experimenter will provide you with more information about the experiment at the completion of this session. If you have any further questions regarding this study, you may email the principal investigator, Hadley Parum (hp4041@bard.edu), or the Bard College Institutional Review Board (irb@bard.edu).

STATEMENT OF CONSENT:
"The purpose of this study, procedures to be followed, and the risks and benefits have been explained to me. I have been told whom to contact if I have additional questions. I have read this consent form and agree to be in this study, with the understanding that I may withdraw at any time."

By continuing to proceed by clicking the below button, you are indicating the following:

- I am at least 18 years of age.

- I agree to be contacted through MTurk with invitations to participate in a future study, should I qualify.
- I have fully read and understand the contents of this informed consent agreement.
- I agree with the above statement of consent.

[ ] I agree with the above statements and provide my informed consent to participate in this study.

Main Experiment Consent Form (MTurk)
INFORMED CONSENT AGREEMENT

You are being asked to participate in a study designed as a part of a Bard College Senior Project in the Department of Psychology. This experiment seeks to investigate how sound and learning interact.

Please take time to thoroughly read through this form, which describes potential risks and benefits of participation in this study. After you have been fully informed, you have the right to choose whether you wish to participate by signing or not signing this form. You also have the ability to end your participation in this study at any time during or after your completion of the study.

Background: This experiment seeks to investigate how sound and learning interact. In this study, you will be asked to pay attention to different visually presented stimuli, including English words and numbers. You will also be asked to listen to various sounds during the learning process. This task will take approximately 25 minutes to complete.

Risks and Benefits: Participating in this study poses minimal risk. As a part of this experiment, you would be asked to complete a series of tasks on the Inquisit Web 6 platform, which is free and safe to download, and compatible with most desktop and mobile devices. You will be asked to listen to different sounds, and will be allowed to determine a comfortable listening volume to avoid potential discomfort. You will also be asked to pay attention to items presented visually through the screen, and will be offered the opportunity to take short breaks periodically to alleviate the potential fatigue or discomfort resulting from time spent in front of a digital device.

While participation in this study may not provide any direct benefits to you, you will receive compensation for your time, described below. Additionally, you will provide crucial information that enhances our understanding of learning and human cognition, and you will support the completion of a Senior Project, therefore supporting an undergraduate student in developing research skills.

Compensation: For your participation, you will be provided with direct monetary compensation through MTurk totalling $3.02 for a 25-minute task.

Confidentiality: Information you provided through the Inquisit platform, including responses to questionnaires and your performance on tasks, will be temporarily stored on Inquisit's servers, and at the conclusion of data collection will be stored offline on password-protect computers to which only the primary investigator and their advisor will have access. These data will be connected to an anonymous subject number unrelated to your MTurk account. The completion code you receive at the completion of the task will be randomly issued by Inquisit and used only to disburse payment, and at no point will Inquisit receive information connected to your MTurk account.

The final published version of this research will be permanently and publicly available as a Senior Project at the Stevenson Library of Bard College, and digitally through the Digital Commons. This information may also be used in the potential publication or presentation on findings that result from this project. In these cases, your data will be presented in aggregate with the data of other participants and will not be linked to any identifiable information.

Your Rights as a Participant: Your participation in this experiment is entirely voluntary, meaning that you can choose to not participate or to withdraw from the experiment at any time without penalty.

The experimenter will provide you with more information about the experiment at the completion of this session. If you have any further questions regarding this study, you may email the principal investigator, Hadley Parum (hp4041@bard.edu), or the Bard College Institutional Review Board (irb@bard.edu).

STATEMENT OF CONSENT:
"The purpose of this study, procedures to be followed, and the risks and benefits have been explained to me. I have been told whom to contact if I have additional questions. I have read this consent form and agree to be in this study, with the understanding that I may withdraw at any time."

By continuing to proceed by clicking the below button, you are indicating the following:

- I am at least 18 years of age.
- I agree to all qualifications (that I have not experienced hearing loss, I do consider English a primary language, am willing to download Inquisit Web 6 and can play audio through my device).
- I have fully read and understand the contents of this informed consent agreement.
- I agree with the above statement of consent.

[ ] I agree with the above statements and provide my informed consent to participate in this study.

APPENDIX C: DEBRIEFING STATEMENTS
Pilot Experiment Debriefing Statement
DEBRIEFING STATEMENT
PLEASE KEEP THIS SHEET FOR YOUR RECORDS

Study Title: Varying Pitch Interval of Polyrhythms
Principal Investigator: Hadley Parum (hp4041@bard.edu)

Thank you for participating in this experiment! This study was designed to investigate the significance of the musical feature of pitch distance to perceptions of musical stimuli. Your participation will help further research about features of musical stimuli and human cognition.

During the procedure, we asked you to complete a lexical decision task, where you judged whether a string of letters was an English word, while listening to different sounds. While we were interested in your performance on this task, it will be used to compare how unconsciously distracting the sounds were, rather than as a score of your individual performance. You were also asked to provide ratings of the sounds you heard. These will be used again to compare a large selection of sounds to each other, so that future research can better ensure that when some of these sounds are used, any effects driven by their differences can best be explained by differences in pitch distance, rather than how familiar or consonant they are to listeners, for example.

Remember that you are free to withdraw consent even after the conclusion of the experiment for any concerns. If you would like to withdraw your data from our analysis, or if you have any further questions or concerns about the research, you are welcome to contact the primary investigator, Hadley Parum (hp4041@bard.edu). If you have concerns about your rights as a research participant, please contact the Bard College IRB at irb@bard.edu.

To extend support during an ongoing pandemic, and in case of any current distress or fatigue, you are encouraged to contact any of the following: Bard Counseling Center (at 845-758-7433), BRAVE (at 1-845-758-7777) or the National Alliance on Mental Illness's (NAMI's) HelpLine (at 1-800-950-6264).

Thank you again for your participation! While this study does not rely on explicit deception concerning its hypotheses, we kindly request that you not disclose the details of the experiment to any individuals until after they have had the opportunity to participate or decline to participate in the study. Prior knowledge of the goals may invalidate the results and weaken our ability to rely on them in the future. We greatly appreciate your cooperation.

Screener Study Debriefing Statement (MTurk)
DEBRIEFING STATEMENT
PLEASE KEEP THIS SHEET FOR YOUR RECORDS

Study Title: Screener Questionnaire for Bard College Senior Project Research
Principal Investigator: Hadley Parum (hp4041@bard.edu)

Thank you for participating in this screener questionnaire! This study was designed to investigate experiences with hearing, music, and language, and will be used to determine eligibility for future research by the principal investigator. Should you be determined to be eligible for future research, you will be contacted directly through MTurk; your contact information beyond this has neither been solicited nor accessed.

Remember that you are free to withdraw consent even after the conclusion of the experiment for any concerns. If you would like to withdraw your data from our analysis or potential participant pool, or if you have any further questions or concerns about the study, you are welcome to contact the primary investigator, Hadley Parum (hp4041@bard.edu). If you have concerns about your rights as a research participant, please contact the Bard College IRB at irb@bard.edu.

To extend support during an ongoing pandemic, and in case of any current distress or fatigue, you are encouraged to contact the following helpline for further support: National Alliance on Mental Illness's (NAMI's) HelpLine (at 1-800-950-6264).

Thank you again for your participation! Please return to MTurk and enter the following unique code in order to finish the HIT and have payment credited to your account.

Code: 248371

Main Experiment Debriefing Statement (MTurk)
DEBRIEFING STATEMENT
PLEASE KEEP THIS SHEET FOR YOUR RECORDS

Study Title: Varying Pitch Interval as Cued Musical Context Demonstrates the Perceptual Similarity of Polyrhythms
Principal Investigator: Hadley Parum (hp4041@bard.edu)

Thank you for participating in this experiment! This study was designed to investigate the relationship between different musical features and memory. Your participation helped progress understanding of how humans may organize or categorize perceptual information and assess the similarity between musical stimuli when behavioral processes depend on this material to some extent.

Remember that you are free to withdraw consent even after the conclusion of the experiment for any concerns. If you would like to withdraw your data from our analysis, or if you have any further questions or concerns about the research, you are welcome to contact the primary investigator, Hadley Parum (hp4041@bard.edu). If you have concerns about your rights as a research participant, please contact the Bard College IRB at irb@bard.edu.

To extend support during an ongoing pandemic, and in case of any current distress or fatigue, you are encouraged to contact the following helpline for further support: National Alliance on Mental Illness's (NAMI's) HelpLine (at 1-800-950-6264).

Thank you again for your participation! While this study does not rely on explicit deception concerning its hypotheses, we kindly request that you not disclose the details of the experiment to any individuals until after they have had the opportunity to participate or decline to participate in the study. Prior knowledge of the goals may invalidate the results and weaken our ability to rely on them in the future. We greatly appreciate your cooperation.

Please return to MTurk and enter the following unique code in order to finish the HIT and have payment credited to your account.

Code: 173842

APPENDIX D: QUESTIONNAIRES

Screening Questionnaire

Instructions: These questions are used only to determine your ability to participate in current research. Your answers will not be connected to your name, unique participant ID, (MTurk account,) or in any other way connected to your identity. (MTurk users only: Due to the different needs of various studies, it is vital that you respond honestly so that we may solicit you for appropriate opportunities.)

Hearing Normality

1. Have you ever been diagnosed with hearing loss (including but not limited to Otosclerosis, Ménière's disease, Presbycusis, Physical head injury, Ototoxic medicine, or noise exposure)?

   [YES] [NO] [DECLINE TO RESPOND] * Must select one

2. Do you currently consider yourself D/deaf or hard of hearing?

   [YES] [NO] [DECLINE TO RESPOND] * Must select one

3. Do you consider yourself to have absolute/perfect pitch?

   [YES] [NO] [DECLINE TO RESPOND] * Must select one

4. Do you consider yourself to be tone deaf?

   [YES] [NO] [DECLINE TO RESPOND] * Must select one

English proficiency

5. Do you consider English a primary language?

   [YES] [NO] [DECLINE TO RESPOND] * Must select one

6. Did your household speak English more than half the time during your childhood?

   [YES] [NO] [DECLINE TO RESPOND] * Must select one

Demographic Questionnaire

1. What is your age in years? *(Provide response in whole numbers, e.g. "3"; NOT "4.5")*

[_________] [Decline to Respond] * Must select one

2. What is your gender identity?

[Male] [Female] [Nonbinary/Other] [Decline to Respond] * Must select one

3. What is your race or ethnicity?

[African-American/Black] [Asian] [Native American] [White/Caucasian]

[Hispanic/Latino] [Other] [Decline to Respond] *Select all that apply

Musical Experience Questionnaire

When asked for a numeric answer, please answer in whole numbers (e.g., "3", "45"; NOT "1.5").

1. Have you ever had musical training (in school, private lessons, etc.)?

   [YES] [NO]

   a. If yes, what was the age at which you began training?

   [_________]

   b. If yes, what was the age at which you stopped playing (current age if you still play)?

   [_________]

2. Do you currently play an instrument (at any level of formality)?

   [YES] [NO]

   a. If yes, for approximately how many hours a week do you play an instrument?

   [_________]

   b. Of these hours, how many hours a week do you spend improvising (rather than playing from memorized or written music)?

   [_________]

3. How many hours per week on average do you spend listening to music of any kind?

   [_________]

Musical Feature Ratings

Think back to the sound that played during the most recent lexical decision task. For each statement below, select the number that best reflects your personal agreement with the statement about the sound you just heard. Select only one number for each statement.

*The sound was **familiar**.*

Strongly Disagree			Neither agree nor disagree			Strongly Agree
1	2	3	4	5	6	7

*The sound was **pleasurable** (felt good to hear).*

Strongly Disagree			Neither agree nor disagree			Strongly Agree
1	2	3	4	5	6	7

*The sound was **consonant** (sounded good aesthetically).*

Strongly Disagree			Neither agree nor disagree			Strongly Agree
1	2	3	4	5	6	7

*The sound was **engaging**.*

Strongly Disagree			Neither agree nor disagree			Strongly Agree
1	2	3	4	5	6	7

*The sound was **distracting**.*

Strongly Disagree			Neither agree nor disagree			Strongly Agree
1	2	3	4	5	6	7

*The sound was **happy** (regardless of what emotion it made you feel).*

Strongly Disagree			Neither agree nor disagree			Strongly Agree
1	2	3	4	5	6	7

APPENDIX E: HUMAN SUBJECTS TRAINING

Certificate of Completion

The National Institutes of Health (NIH) Office of Extramural Research certifies that **Hadley Parum** successfully completed the NIH Web-based training course "Protecting Human Research Participants".

Date of completion: 01/21/2018.

Certification Number: 2581367.

IRB Proposal, Pitch Interval Context and Perceptual Similarity of Polyrhythms 21

APPENDIX F: MUSICAL STIMULI

Audio Examples

For an example of the looped audio a participant would hear concurrently with the lexical decision (pilot) or list learning (main experiment) task, click here or type bit.ly/5-4-loop into your browser.

Musical notation for full stimuli

IRB Proposal, Pitch Interval Context and Perceptual Similarity of Polyrhythms 22

Entry C2. IRB Approval, November 15, 2020.

Bard College   Institutional Review Board

Date: November 15, 2020
To: Hadley Parum
Cc: Justin Hulbert, Deborah Treadway, Brandt Burgess
From: Tom Hutcheon, IRB Chair
Re: Varying Pitch Interval as Cued Musical Context Demonstrates the Perceptual Similarity of Polyrhythms

DECISION: APPROVED

Dear Hadley,

The Bard Institutional Review Board has reviewed and approved your proposal entitled “Varying Pitch Interval as Cued Musical Context Demonstrates the Perceptual Similarity of Polyrhythms.” Your proposal is approved through November 15, 2021 and your case number is: 2020NOV15-PAR.

Please notify the IRB if your methodology changes or unexpected events arise.

This sounds like a really interesting project and we wish you the best of luck with your research!

Tom Hutcheon
IRB Chair
thutcheo@bard.edu

Entry C3. IRB Amendment, submitted January 3, 2021.

Dear Members of the IRB,

Attached you will find an amended version of the IRB protocol 2020NOV15-PAR ("Varying Pitch Interval as Cued Musical Context Demonstrates the Perceptual Similarity of Polyrhythms"). While all changes from the original version of the submission are highlighted in yellow so for ease of reading, I will also describe the changes and their motivation below.

During December 2020, I was lucky enough to receive monetary support for this project both from the Psi Chi International Honor Society and Psychology, as well as from Bard's Lifetime Learning Institute. Additionally, completion of the coding and testing process brought to light that we could tweak the number of intervals included (now 36 instead of 37, without the unison), increase how many of each participants hear (12 rather than 10), allowing for more robust observations while also decreasing the overall time needed to be spent on the task (30 minutes instead of 45).

As a result of both the additional financial support and the streamlined pilot protocol, I would be able to increase the number of participants I am able to recruit threefold, from 24 to 72. Additionally, the new time estimates are reflected in an adjusted payment amount for those who participate. The monetary support received also accounts for the increase in participants I would be able to retain for the main experiment, which is now 559 instead of the previous 390.

The ability to increase the number of participants included in the pilot experiment is especially exciting, because having a more robust understanding of how different potential musical stimuli affect participants' performance on a word task as well as what opinions people form of these sounds will greatly improve the ability of the main experiment to account for these differences in design and interpretation. The increase in number improves the likelihood that any findings are not the result of chance alone, and an increase in the diversity of participants in the pilot would additionally contribute resilience of any findings when applied to a more general population, such as that which will be included in the main experiment.

My initial IRB submission included language about recruiting "members of the Bard community ... through publicity measures such as posters in campus locations or through social media." Additionally, I would like to include social media advertisement that reaches potentially beyond the Bard community, to other people with access to Facebook, Instagram, and/or Twitter. This may include college-aged individuals as well as older individuals who have interacted with me in professional or familial settings, or who have had involvement in the debate community, for example. If these forms of advertisement do not garner the intended number of participants, recruitment may also extend to workers MTurk, providing access to a population similar to those who would be involved in the main experiment (in order to minimize selection differences between MTurk workers involved in the experiment rather than pilot, all MTurk workers' compensation will remain commensurate with the U.S. Federal minimum wage).

This intended change to accessed populations comes with changes to recruitment materials in Appendix A, and the inclusion of a global test of eligibility criteria when informed consent is administered. This global test of eligibility criteria allows for no time delay between administration and completion of the main task (this time delay exists for the main experiment), and allows for non-MTurk participants to provide a holistic response to their eligibility without specific answers being linked to payment information they provide; no comparable risk of identification exists for MTurk participants, but the structure is retained in order to minimize variation in the overall task. This adjustment of the screening questionnaire is reflected in Appendix D. Changes to the pilot consent form in Appendix B reflect the

aforementioned changes to payment, task length, and exclusion criteria. Finally, an outline of the Qualtrics survey to which non-MTurk pilot participants will be directed is included in Appendix G.

Ultimately, since these changes do not affect the content of the experiment yet do maintain attention to participants' security and compensation commensurate with task length, there are no expected changes to the risks or benefits associated with the experiment. However, these changes would allow for great improvement to the quality of data collected both in the pilot study, and in the main experiment which will rely on its findings.

Of course, let me know if I can provide any additional information or clarify anything within.

Sincerely,
Hadley Parum
hp4041@bard.edu

P.S., Happy New Year!

Section 1

1. **Today's date:** January 3, 2021
2. **Name:** Hadley Parum
3. **Email:** hp4041@bard.edu
4. **Your Academic Program/Department/Office:** Psychology Department
5. **Your status (faculty, staff, graduate or undergraduate student):** Undergraduate student, Senior I
6. **Adviser or Faculty Sponsor (if applicable):** Justin Hulbert
7. **If you are a graduate or undergraduate student, has your Adviser or Faculty Sponsor seen and approved your application?** Yes
8. **Your Adviser's or Faculty Sponsor's email address (if applicable):** jhulbert@bard.edu
9. **Please list all individuals (full name and status, i.e. faculty, staff, student) involved in this project that will be working with human subjects. Note: Everyone listed must have completed Human Subject Research Training within the past three years.** Hadley Parum
10. **Do you have external funding for this research?** Yes.
11. **If so, state the name of the sponsor and the title of the project as it was submitted to that sponsor.** I received a Fall Undergraduate Research Grant amounting $1,500 from the Psi Chi Honor Society, as well as a Seniors-to-Seniors Grant from the Lifetime Learning Institute amounting to $750. Both were received in December 2020, and the title of this project as submitted to both was "Varying Pitch Interval as Cued Musical Context Demonstrates the Perceptual Similarity of Polyrhythms"

Section 2

1. **What is the title of your project?** Varying Pitch Interval as Cued Musical Context Demonstrates the Perceptual Similarity of Polyrhythms
2. **When do you plan to begin this project? (Start date):** December 1, 2020
3. **Describe your research question(s):** We know that pitch and rhythm interact to explain our perception of music. However, it is not fully understood what aspects of these key features are most essential to the perceptual similarity of musical examples. To investigate this further, the present research relies on a test of context-dependent memory of English words using a test of delayed free recall. Participants will learn words while hearing a polyrhythm played at the pitch interval of a perfect fifth, and will later be tested on the studied words while hearing the polyrhythm at the same or a varied pitch interval (e.g., the minor second, major third, and major sixth), or will hear no additional sound. It is expected that participants will recall a higher proportion of words correctly if the musical-auditory context is the same at test as it was during learning. Furthermore, I expect recall to be higher if participants heard a) pitch intervals in the same octave level as that heard at learning, such that intervals are closer rather than farther apart, and if they heard b) an interval class consonant to the original interval or the ratio of the polyrhythm. These findings aim to show that perceptual relationships like similarity depend on certain relational aspects of key features, like proximity and consonance, furthering understanding of perception and music cognition.
4. **Describe the population(s) you plan to recruit and how you plan to recruit participants. Please submit all recruitment material, emails and scripts to IRB@bard.edu.** Participants will be recruited from a mix of sources, including the Bard community, additional online

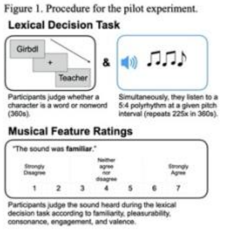
populations reached through social media, and through Amazon's Mechanical Turk (MTurk). Before they agree to participate, they will be determined to be above the age of 18, with English as a dominant language, and without self-reported hearing impairments or perfect pitch; MTurk participants will also be restricted to those currently residing in the United States. Additionally, all participants will verify that the device used for the experiment has functional audio output and the capacity to download the Inquisit Web 6 player,[1] which is compatible with most Windows, Mac, iOS, and Android devices. All participants will be informed that the experiment will involve listening to audio and downloading the Inquisit program before required to begin the experiment. Any participants recruited through MTurk will be provided a description of the task's purpose, length, and use of audio playback as well as the Inquisit Web player, and will be able to volunteer through Amazon's website to participate in exchange for disclosed monetary compensation. Any participants recruited from the Bard community may be solicited through publicity measures such as posters in campus locations or through social media. Additionally, social media advertisements may reach participants from beyond the Bard community, including participants with access to Facebook, Instagram, and/or Twitter.

5. **Will your participants include individuals from vulnerable or protected populations (e.g., children, pregnant women, prisoners, or the cognitively impaired)?** No.
6. **If your participants will include individuals from the above populations, please specify the population(s) and describe any special precautions you will use to recruit and consent.** N/A.
7. **Approximately how many individuals do you expect to participate in your study?** Total estimate: 852unique participants. Pilot study: 72 participants; screener questionnaire: 780 participants; main experiment: 559 participants.
8. **Describe the procedures you will be using to conduct your research. Include descriptions of what tasks your participants will be asked to do, and about how much time will be expected of each individual. NOTE: If you have supporting materials (printed surveys, questionnaires, interview questions, etc.), email these documents separately as attachments to IRB@bard.edu. Name your attachments with your last name and a brief description (e.g., "WatsonSurvey.doc).**
   a. Pilot Study
      i. Summary: A pilot study will be conducted to measure how potential musical stimuli vary on extra-musical characteristics, in order to better inform the stimuli used during the main experiment. Figure 1 provides a visual overview of the procedures for this 25-30 minute task.. Participants recruited through the Bard Community or otherwise reached through social media will be directed to respond to global exclusion criteria in a Qualtrics survey (see Appendix D), and those who qualify will then be directed to complete the main task in the Inquisit Web 6 platform. Participants recruited through MTurk will complete the global exclusion criteria as a part of the informed consent measure (see Appendix B) ~~The time window during which the participant takes the study will be scheduled with the primary investigator, but the scheduling and completion of the experiment will not require the participant to physically meet the primary investigator at any point.~~ The participant will complete informed consent

[1] View download page and compatibility statement: https://www.millisecond.com/download/inquisitweb6.aspx

measure (see Appendix B), followed by a short interval recognition task, described below. Then, participants will repeat a short listening procedure twelve times. This listening procedure consists of completing the lexical decision task, described below while hearing a given musical stimuli on loop, and then in silence completing the musical feature ratings for the given audio (see Appendix D), where they answer Likert-type questions expressing dis/agreement with statements such as "The sound was familiar." Following that approximately 30 minute listening procedure, they will complete short demographic and musical experience questionnaires (see Appendix D) and then directed to return to the Qualtrics survey (for non-MTurk participants) or to the MTurk page (for MTurk participants) with a completion code. There, they will be thanked, debriefed, and will receive payment.

Figure 1. Procedure for the pilot experiment.



ii. *Interval Recognition Task.* In order to confirm participant self-reports about hearing loss, perfect pitch and pitch deafness, and working audio equipment, they will complete a short (less than one minute) interval recognition task. Six musical examples, consisting of two notes played in sequence at 150bpm (for a five second musical example), will be presented to participants. After each one, they will be asked whether the second note was higher or lower than the first. Then, they will be asked to guess which interval name best fits the interval they heard between the two notes (e.g., minor second, perfect fourth, or major sixth).

iii. *Musical Stimuli.* Created with Musescore, sounds will consist of 5:4 polyrhythms played at a given pitch interval, ranging from a minor second to three octaves apart, for 36 total possible intervals included in the pilot experiment. Based on pilot experiment results, 12 of these intervals will be included in the main experiment. See Appendix F for examples of musical stimuli in music notation and for links to audio examples.

iv. *Lexical Decision Task.* Participants will complete a task adapted from that which is publicly available on Inquisit's millisecond database.[2] in order to provide a within-subject measure of attention while evaluating words. They will be asked to categorize a given set of letter characters as either comprising an English word or a nonword, responding with an "I" key for valid words and an "E" for nonwords. Each word is presented briefly, for a period of 250ms, followed by a fixation cross in the center of the screen until their response, which is followed by 700ms of buffer time in

[2] View the online list of publicly available tasks: https://www.millisecond.com/download/library/lexicaldecisiontask/

which only the fixation cross is present. Including short test trials to acclimate participants to the controls, the presentation of all words takes approximately three minutes.

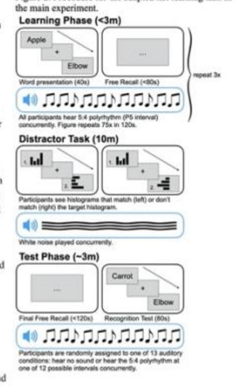
b. Screener Questionnaire

i. Summary: A screener questionnaire is a procedure supported by the community of researchers and MTurk workers in order to equitably determine eligible participants for future research, without disqualifying MTurk workers who would otherwise attempt to complete a study for which they are ineligible. Participants recruited through MTurk who complete the informed consent agreement will then complete the screening questionnaire and the demographics questionnaire (see Appendix D) in the Inquisit Web 6 platform. Following this, they will be thanked, debriefed, and directed to return to MTurk with a code to ensure payment. The whole procedure will take approximately two minutes to complete.

c. Main Experiment

i. Summary: Participants on MTurk who completed the Screening task and are determined to qualify will be solicited directly through MTurk's contact system and invited to participate in an approximately 25-minute task. Figure 2 provides a visual overview of the procedures discussed here. Those who join and complete the informed consent agreement will then complete the interval recognition task (described above, under the pilot experiment). Following this, all participants will complete the learning portion of the List Learning Task, described below, while listening to the same musical stimuli (the 5:4 polyrhythm at a predetermined consonant interval, e.g. the perfect fifth). Participants will then listen to white noise for the duration of a ten minute visual distractor task. Afterwards, participants will complete the final tests of the List Learning Task while randomly assigned to one of 13 conditions, in which they may hear no audio, or one of the 12 musical stimuli included in the main experiment. Finally, participants will complete demographics and musical experience questionnaires (see Appendix D), and then will be thanked, debriefed, and directed to return to MTurk with a code to ensure payment.

Figure 2. Procedure for the adapted list learning task in the main experiment.



ii. *List Learning Task.* In order to measure delayed free recall and recognition of learned material, the Inquisit List Learning Task[3] will be adapted for present use in Inquisit 6 Web (2020). Participants will learn and be tested on the same 20 unrelated, neutrally valenced English words for objects (e.g., "elbow," "apple"). The task consists of a learning phase, in which the words are presented in fixed order every two seconds, followed by a short free recall period without feedback, three times. After a ten minute break or distractor task, there is a test of delayed recall lasting two minutes, where participants are asked to recall as many words as they can, in any order. Finally, to test delayed recognition, 20 words will be presented and participants will use key responses to report whether the word was in the studied list ("Y") or not ("N"). Half of the final 20 words were on the original list of 20 words, and the other half are 10 unstudied words.

iii. *Distractor Task.* Participants will complete a 10-minute visuospatial reasoning task, in order to provide a break between learning and test that occupies attentional resources, both to engage participants and limit rehearsal. To this end, participants will complete a task such as the Manikin Test[4] of Spatial awareness, in which participants must make judgements about which hand of a figure holds the same shape as that which surrounds the figure, or the Spatial Processing Task[5], in which participants must judge whether a histogram is a transformed version of a previously presented target histogram, or a differently shaped histogram.

9. **Describe any risks and/or benefits your research may have for your participants.**
   a. Risks: All components of this study involve minimal risk for participants. Due to the online nature of all tasks, there exists a small risk for discomfort or fatigue during stimulus delivery. For those involved in the pilot study and main experiment, the listening tasks may pose the additional small risk of auditory discomfort. To alleviate these risks, participants will a) be told in advance about the occurrence and length of exposure to these stimuli, b) be able to set the volume to a comfortable level before listening tasks, and c) be offered short, self-directed breaks at various points of the procedures. Since listening to audio and using electronic devices is a frequent occurrence for most participants, it is likely that they will be familiar with their comfort level during such tasks and be able to make informed decisions about participation and breaks.
   b. Benefits: This study offers no direct benefits to participants. They may receive indirect emotional benefits as a result of the knowledge that they have contributed towards scientific inquiry and the completion of a Senior Project. All participants will receive compensation for their time, according to applicable minimum hourly wages at the time of their participation:

[3] View the List Learning Task's code, user manual, and demo: https://www.millisecond.com/download/library/listlearningtask/
[4] View Manikin Test of Spatial Orientation and Transformations code, user manual, and demo: https://www.millisecond.com/download/library/mentalrotation/manikintest/
[5] View the Spatial Processing Task's code, user manual, and demo: https://www.millisecond.com/download/library/mentalrotation/spatialprocessingtask/

i. Participants recruited from the Bard Community or otherwise through social media measures for the pilot study will receive $6.25 (for a 30-minute task and a New York State minimum wage of $12.50/hour as of January 1, 2021).

ii. Participants recruited from MTurk for the pilot study will receive $3.75 (for a 30-minute task at the U.S. Federal minimum wage of $7.25/hour).

iii. Participants recruited from MTurk for the screening questionnaire will receive $0.24 (for a 2-minute task at the U.S. Federal minimum wage of $7.25/hour).

iv. Participants recruited from MTurk for the main experiment will receive $3.02 (for a 25-minute task at the U.S. Federal minimum wage of $7.25/hour).

10. **Describe how you plan to mitigate (if possible) any risks the participants may encounter.** Since listening to audio and using electronic devices is a frequent occurrence for most participants, it is likely that they will be familiar with their comfort level during such tasks. Since listening to audio and using electronic devices is a frequent occurrence for most participants, it is likely that they will be familiar with their comfort level during such tasks. In order to further defray any potential discomfort, participants will:
    a. Be told in advance about the occurrence and length of exposure to these visual and auditory stimuli;
    b. Be able to set the volume to a comfortable level during the screening procedure, and
    c. Be offered short, self-directed breaks at various points of the procedures.
11. **Describe the consent process (i.e., how you will explain the consent form and the consent process to your participants):** All participants will be presented with the relevant consent form prior to the delivery of the task through Inquisit. The form contains general information about the goals of the experiment, risks and benefits, compensation, and confidentiality details, as well as the contact information for the primary investigator and the Bard Institutional Review Board.
12. **Have you prepared a consent form(s) and emailed it as an attachment to IRB@bard.edu?** Yes.
13. **If you are collecting data via media capture (video, audio, photos), have you included a section requesting consent for this procedure(s) in your consent form(s)?** N/A.
14. **If your project will require you to employ a verbal consent process (no written consent forms), please describe why this process is necessary and how verbal consent will be obtained and stored.** N/A.
15. **What procedures will you use to ensure that the information your participants provide will remain confidential and safeguarded against improper access or dissemination?** Participants will be assured during the consent and debriefing processes that steps are being taken to ensure the safeguarding of any personal information. Information collected through the Inquisit Web platform (answers to questionnaires and data from experimental tasks) will not be connected to their identifying information (for in-person participants, the name on their consent form; for MTurk participants, any information associated with their MTurk account) after the conclusion of data collection. All information provided through the Inquisit Web platform will be temporarily stored on their servers in Oregon, USA. Their security statement is accessible here. While it is stored there, it is only accessible to the primary investigator through their password-protected Inquisit account. At the conclusion of data collection, data will be stored offline on a password-protected computer to which only the primary investigator has access. Potentially

Pitch Interval as Context (2020NOV15-PAR, Amendment 1) 9

personal information collected (e.g., responses to the screening questionnaire) will only be used to determine qualification for further participation, and not connected to their scores in the main experiment, and all data (including demographics and other questionnaires, and task performance) will be reported in aggregate.

16. **Will it be necessary to use deception with your participants at any time during this research? Withholding details about the specifics of one's hypothesis does not constitute deception, this is called incomplete disclosure. Deception involves purposefully misleading participants about the nature of the research question or about the nature of the task they will be completing.** No.
17. **If your project study includes deception, please describe here the process you will use, why the deception is necessary, and a full description of your debriefing procedures.** N/A.
18. **For all projects, please include your debriefing statement. (This is information you provide to the participant at the end of your study to explain your research question more fully than you may have been able to do at the beginning of the study.) All studies must include a debriefing statement. Be sure to give participants the opportunity to ask any additional questions they may have about the study.** See Appendix C. for the debriefing statements.
19. **If you will be conducting interviews in a language other than English, will you conduct all of the interviews yourself, or will you have the assistance of a translator? If you will be using the assistance of a translator, that individual must also certify that he or she is familiar with the human subject protocol and has completed the online training course.** N/A.
20. **If your recruitment materials or consent forms will be presented in languages other than English, please translate these documents and email copies to IRB@bard.edu.** N/A.

Entry C4. IRB Approval, January 12, 2021.

Bard College Institutional Review Board

Date: January 12, 2021
To: Hadley Parum
Cc: Justin Hulbert, Deborah Treadway, Brandt Burgess
From: Tom Hutcheon, IRB Chair
Re: Proposed Amendments to 2020NOV15-PAR

DECISION: APPROVED

Dear Hadley,

The Bard Institutional Review Board has reviewed and approved the amendments you submitted to your protocol on January 3, 2021. Your case number remains 2020NOV15-PAR.

Please notify the IRB if your methodology changes or unexpected events arise.

This sounds like a really interesting project and we wish you the best of luck with your research!

Tom Hutcheon
IRB Chair
thutcheo@bard.edu

Entry C5. IRB Amendment, submitted March 8, 2021.

Dear Members of the IRB,

Attached you will find an amended version of the IRB protocol 2020NOV15-PAR ("Varying Pitch Interval as Cued Musical Context Demonstrates the Perceptual Similarity of Polyrhythms"). The changes to the protocol itself are highlighted in yellow so for ease of reading. They consist of (a) an updated date and title in Section 1, (b) a small adjustment to experimental methodology in Section 2, question 8, and (c) a corresponding adjustment to compensation as discussed in Section 2, question 9 and experimental recruitment and consent materials. .

The IRB protocol originally submitted and maintained through the first amendment featured an experimental procedure testing context-dependent memory, with participants to study words for an immediate test of recall, and later a delayed test of recall. This method has participants intentionally study words, and is beneficial for the non-associative form of learning that may be encouraged in participants. In a meta-analysis of context-dependent memory studies involving the characteristics of surrounding rooms, Smith and Vela (2001) found that effect sizes for context-dependent memory effects were significantly greater in studies with non-associative, rather than associative processing.

However, only one published study has tested *music*-dependent memory using intentionally learned words, and its final test of recall was delivered after a 48-hour delay (Smith, 1985). My study intends to employ a final test after a delay shorter than five minutes. Work on music-dependent memory has not produced evidence for a music-dependent memory effect for intentionally learned words when tested in this short timeframe, with Isarida et al. (2008) instead finding a null effect in this case. However, the authors did find significant effects of background music context when participants learned words *incidentally*, when the learning phase consisted of participants rating words for their pleasure on a likert-type scale. This is consistent with many other recent studies in music-dependent memory, which utilize incidental learning, short delays between learning and test, and have produced significant effects of musical, genre, and tempo as contexts (Balch et al., 1992; Balch & Lewis, 1996; Isarida et al., 2017; Mead & Ball, 2007).

In order to more closely replicate the methodologies of more recent work in music-dependent memory and employ a paradigm in which there already exists evidence that context changes can differentially affect recall, I have amended this IRB protocol to also allow for the learning phase in my study to either ask participants to "study the list [of 20 words] for immediate tests of recall, or will be asked to rate the pleasantness of words on a scale of 1-5, where 5 indicates a very pleasant word" (p. 6). Additionally, since "words will be presented twice, in random order" (p. 6), the rating task will take fewer than 240 seconds, with each word being presented for a maximum of 5 seconds and a 1 second gap between words. Accordingly, the distraction task will no longer last for 10 minutes, but "after practice, will last 200 seconds" (p. 6). This way, the word rating and visuospatial tasks will last for approximately the same amount of time; additionally, a 4-minute delay between incidental learning and recall is within a 30 second–5 minute range commonly employed by music-dependent memories cited above.

These changes would not greatly change the risks and benefits experienced by participants in the study. The overall task length does decrease with a shorter distraction task, such that participants will be compensated $2.50 for a 18-20 minute task, rather than $3.02 for an approximately 25-minute task (p. 8, 11, 15-16). Thus, compensation still remains slightly above the U.S. Federal minimum wage, as it stands currently. Additionally, the overall time in which participants need to be engaged in the task and exposed

to experimental sounds or white noise has decreased, which may diminish risks associated with electronic exposure or sensory fatigue.

I hope that these changes to the study may strengthen the weight of any findings produced, and am happy that doing so does not bring novel risks to participants. As always, let me know if I can provide additional information.

Sincerely,
Hadley Parum
hp4041@bard.edu

References:

Balch, W. R., Bowman, K., & Mohler, L. A. (1992). Music-dependent memory in immediate and delayed word recall. *Memory & Cognition*, *20*(1), 21–28. https://doi.org/10.3758/BF03208250

Balch, W. R., & Lewis, B. S. (1996). Music-Dependent Memory: The Roles of Tempo Change and Mood Mediation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(6), 1354–1363.

Isarida T. K., Isarida T., & Hayashibe K. (2008). *Context dependent effects of background music on the free recall of incidentally and intentionally learned words.* https://doi.org/10.5265/JCOGPSY.5.107

Isarida, T. K., Kubota, T., Nakajima, S., & Isarida, T. (2017). Reexamination of Mood-Mediation Hypothesis of Background-Music-Dependent Effects in Free Recall. *Quarterly Journal of Experimental Psychology*, *70*(3), 533–543. https://doi.org/10.1080/17470218.2016.1138975

Mead, K., & Ball, L. (2007). Music tonality and context-dependent recall: The influence of key change and mood mediation. *European Journal of Cognitive Psychology - EUR J COGN PSYCHOL*, *19*, 59–79. https://doi.org/10.1080/09541440600591999

Smith, S. M. (1985). Background Music and Context-Dependent Memory. *The American Journal of Psychology*, *98*(4), 591–603. JSTOR. https://doi.org/10.2307/1422512

Smith, S. M., & Vela, E. (2001). Environmental context-dependent memory: A review and meta-analysis. *Psychonomic Bulletin & Review*, *8*(2), 203–220. https://doi.org/10.3758/BF03196157

Section 1

1. **Today's date:** March 8, 2021
2. **Name:** Hadley Parum
3. **Email:** hp4041@bard.edu
4. **Your Academic Program/Department/Office:** Psychology Department
5. **Your status (faculty, staff, graduate or undergraduate student):** Undergraduate student, Senior I
6. **Adviser or Faculty Sponsor (if applicable):** Justin Hulbert
7. **If you are a graduate or undergraduate student, has your Adviser or Faculty Sponsor seen and approved your application?** Yes
8. **Your Adviser's or Faculty Sponsor's email address (if applicable):** jhulbert@bard.edu
9. **Please list all individuals (full name and status, i.e. faculty, staff, student) involved in this project that will be working with human subjects. Note: Everyone listed must have completed Human Subject Research Training within the past three years.** Hadley Parum
10. **Do you have external funding for this research?** Yes.
11. **If so, state the name of the sponsor and the title of the project as it was submitted to that sponsor.** I received a Fall Undergraduate Research Grant amounting $1,500 from the Psi Chi Honor Society, as well as a Seniors-to-Seniors Grant from the Lifetime Learning Institute amounting to $750. Both were received in December 2020, and the title of this project as submitted to both was "Varying Pitch Interval as Cued Musical Context Demonstrates the Perceptual Similarity of Polyrhythms"

Section 2

1. **What is the title of your project?** The role of a polyrhythm's pitch interval in music-dependent memory
2. **When do you plan to begin this project? (Start date):** December 1, 2020
3. **Describe your research question(s):** We know that pitch and rhythm interact to explain our perception of music. However, it is not fully understood what aspects of these key features are most essential to the perceptual similarity of musical examples. To investigate this further, the present research relies on a test of context-dependent memory of English words using a test of delayed free recall. Participants will learn words while hearing a polyrhythm played at the pitch interval of a perfect fifth, and will later be tested on the studied words while hearing the polyrhythm at the same or a varied pitch interval (e.g., the minor second, major third, and major sixth), or will hear no additional sound. It is expected that participants will recall a higher proportion of words correctly if the musical-auditory context is the same at test as it was during learning. Furthermore, I expect recall to be higher if participants heard a) pitch intervals in the same octave level as that heard at learning, such that intervals are closer rather than farther apart, and if they heard b) an interval class consonant to the original interval or the ratio of the polyrhythm. These findings aim to show that perceptual relationships like similarity depend on certain relational aspects of key features, like proximity and consonance, furthering understanding of perception and music cognition.
4. **Describe the population(s) you plan to recruit and how you plan to recruit participants. Please submit all recruitment material, emails and scripts to IRB@bard.edu.** Participants will be recruited from a mix of sources, including the Bard community, additional online

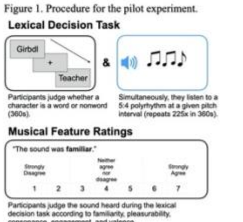
populations reached through social media, and through Amazon's Mechanical Turk (MTurk). Before they agree to participate, they will be determined to be above the age of 18, with English as a dominant language, and without self-reported hearing impairments or perfect pitch; MTurk participants will also be restricted to those currently residing in the United States. Additionally, all participants will verify that the device used for the experiment has functional audio output and the capacity to download the Inquisit Web 6 player,[1] which is compatible with most Windows, Mac, iOS, and Android devices. All participants will be informed that the experiment will involve listening to audio and downloading the Inquisit program before required to begin the experiment. Any participants recruited through MTurk will be provided a description of the task's purpose, length, and use of audio playback as well as the Inquisit Web player, and will be able to volunteer through Amazon's website to participate in exchange for disclosed monetary compensation. Any participants recruited from the Bard community may be solicited through publicity measures such as posters in campus locations or through social media. Additionally, social media advertisements may reach participants from beyond the Bard community, including participants with access to Facebook, Instagram, and/or Twitter.

5. **Will your participants include individuals from vulnerable or protected populations (e.g., children, pregnant women, prisoners, or the cognitively impaired)?** No.
6. **If your participants will include individuals from the above populations, please specify the population(s) and describe any special precautions you will use to recruit and consent.** N/A.
7. **Approximately how many individuals do you expect to participate in your study?** Total estimate: 852 unique participants. Pilot study: 72 participants; screener questionnaire: 780 participants; main experiment: 559 participants.
8. **Describe the procedures you will be using to conduct your research. Include descriptions of what tasks your participants will be asked to do, and about how much time will be expected of each individual. NOTE: If you have supporting materials (printed surveys, questionnaires, interview questions, etc.), email these documents separately as attachments to IRB@bard.edu. Name your attachments with your last name and a brief description (e.g., "WatsonSurvey.doc).**
   a. Pilot Study
      i. Summary: A pilot study will be conducted to measure how potential musical stimuli vary on extra-musical characteristics, in order to better inform the stimuli used during the main experiment. Figure 1 provides a visual overview of the procedures for this 25-30 minute task.. Participants recruited through the Bard Community or otherwise reached through social media will be directed to respond to global exclusion criteria in a Qualtrics survey (see Appendix D), and those who qualify will then be directed to complete the main task in the Inquisit Web 6 platform. Participants recruited through MTurk will complete the global exclusion criteria as a part of the informed consent measure (see Appendix B). The participant will complete informed consent measure (see Appendix B), followed by a short interval recognition task, described below. Then, participants will repeat a short listening procedure twelve times. This listening procedure consists of completing the lexical decision task, described below while

[1] View download page and compatibility statement: https://www.millisecond.com/download/inquisitweb6.aspx

hearing a given musical stimuli on loop, and then in silence completing the musical feature ratings for the given audio (see Appendix D), where they answer Likert-type questions expressing dis/agreement with statements such as "The sound was familiar." Following that approximately 30 minute listening procedure, they will complete short demographic and musical experience questionnaires (see Appendix D) and then directed to return to the Qualtrics survey (for non-MTurk participants) or to the MTurk page (for MTurk participants) with a completion code. There, they will be thanked, debriefed, and will receive payment.

Figure 1. Procedure for the pilot experiment.



ii. *Interval Recognition Task.* In order to confirm participant self-reports about hearing loss, perfect pitch and pitch deafness, and working audio equipment, they will complete a short (less than one minute) interval recognition task. Six musical examples, consisting of two notes played in sequence at 150bpm (for a five second musical example), will be presented to participants. After each one, they will be asked whether the second note was higher or lower than the first. Then, they will be asked to guess which interval name best fits the interval they heard between the two notes (e.g., minor second, perfect fourth, or major sixth).

iii. *Musical Stimuli.* Created with Musescore, sounds will consist of 5:4 polyrhythms played at a given pitch interval, ranging from a minor second to three octaves apart, for 36 total possible intervals included in the pilot experiment. Based on pilot experiment results, 12 of these intervals will be included in the main experiment. See Appendix F for examples of musical stimuli in music notation and for links to audio examples.

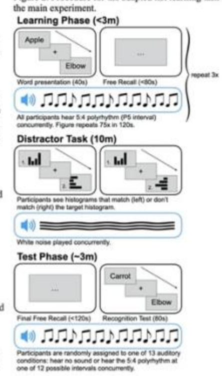
iv. *Lexical Decision Task.* Participants will complete a task adapted from that which is publicly available on Inquisit's millisecond database.[2] in order to provide a within-subject measure of attention while evaluating words. They will be asked to categorize a given set of letter characters as either comprising an English word or a nonword, responding with an "I" key for valid words and an "E" for nonwords. Each word is presented briefly, for a period of 250ms, followed by a fixation cross in the center of the screen until their response, which is followed by 700ms of buffer time in which only the fixation cross is present. Including short test trials to acclimate participants to the controls, the presentation of all words takes approximately three minutes.

[2] View the online list of publicly available tasks: https://www.millisecond.com/download/library/lexicaldecisiontask/

b. Screener Questionnaire
   i. Summary: A screener questionnaire is a procedure supported by the community of researchers and MTurk workers in order to equitably determine eligible participants for future research, without disqualifying MTurk workers who would otherwise attempt to complete a study for which they are ineligible. Participants recruited through MTurk who complete the informed consent agreement will then complete the screening questionnaire and the demographics questionnaire (see Appendix D) in the Inquisit Web 6 platform. Following this, they will be thanked, debriefed, and directed to return to MTurk with a code to ensure payment. The whole procedure will take approximately two minutes to complete.

c. Main Experiment
   i. Summary: Participants on MTurk who completed the Screening task and are determined to qualify will be solicited directly through MTurk's contact system and invited to participate in an approximately 25-minute task. Figure 2 provides a visual overview of the procedures discussed here. Those who join and complete the informed consent agreement will then complete the interval recognition task (described above, under the pilot experiment). Following this, all participants will learn a list of twenty words while listening to the same musical stimuli (the 5:4 polyrhythm at a predetermined consonant interval, e.g. the perfect fifth). Participants will then listen to white noise for the duration of a ten minute visual distractor task. Afterwards, participants will complete the final tests of the List Learning Task while randomly assigned to one of 13 conditions, in which they may hear no audio, or one of the 12 musical stimuli included in the main experiment. Finally, participants will complete demographics and musical experience questionnaires (see Appendix D), and then will be thanked, debriefed, and directed to return to MTurk with a code to ensure payment.
   ii. *List Learning Task.* In order to measure delayed free recall and recognition of learned material, the Inquisit List Learning Task[3] will be adapted for present use in Inquisit 6

Figure 2. Procedure for the adapted list learning task in the main experiment.



[3] View the List Learning Task's code, user manual, and demo: https://www.millisecond.com/download/library/listlearningtask/

Web (2020). Participants will learn and be tested on the same 20 unrelated, neutrally valenced English words for objects (e.g., "elbow," "apple"). During the learning phase, participants may either be asked to study the list for immediate tests of recall, or will be asked to rate the pleasantness of words on a scale of 1-5, where 5 indicates a very pleasant word. Words will be presented twice, in random order. ~~The task consists of a learning phase, in which the words are presented in fixed order every two seconds, followed by a short free recall period without feedback, three times.~~ After a ~~ten minute break or~~ distractor task, there is a test of delayed recall lasting two minutes, where participants are asked to recall as many words as they can, in any order. Finally, to test delayed recognition, 20 words will be presented and participants will use key responses to report whether the word was in the studied list ("Y") or not ("N"). Half of the final 20 words were on the original list of 20 words, and the other half are 10 unstudied words.

iii. *Distractor Task.* Participants will complete a ~~10-minute~~ visuospatial reasoning task, in order to provide a break between learning and test that occupies attentional resources, both to engage participants and limit rehearsal. To this end, participants will complete a task such as the Manikin Test[4] of Spatial awareness, in which participants must make judgements about which hand of a figure holds the same shape as that which surrounds the figure, or the Spatial Processing Task[5], in which participants must judge whether a histogram is a transformed version of a previously presented target histogram, or a differently shaped histogram. The main portion of the distraction task, after practice, will last 200 seconds.

9. **Describe any risks and/or benefits your research may have for your participants.**
   a. Risks: All components of this study involve minimal risk for participants. Due to the online nature of all tasks, there exists a small risk for discomfort or fatigue during stimulus delivery. For those involved in the pilot study and main experiment, the listening tasks may pose the additional small risk of auditory discomfort. To alleviate these risks, participants will a) be told in advance about the occurrence and length of exposure to these stimuli, b) be able to set the volume to a comfortable level before listening tasks, and c) be offered short, self-directed breaks at various points of the procedures. Since listening to audio and using electronic devices is a frequent occurrence for most participants, it is likely that they will be familiar with their comfort level during such tasks and be able to make informed decisions about participation and breaks.
   b. Benefits: This study offers no direct benefits to participants. They may receive indirect emotional benefits as a result of the knowledge that they have contributed towards scientific inquiry and the completion of a Senior Project. All participants will receive compensation for their time, according to applicable minimum hourly wages at the time of their participation:

[4] View Manikin Test of Spatial Orientation and Transformations code, user manual, and demo: https://www.millisecond.com/download/library/mentalrotation/manikintest/
[5] View the Spatial Processing Task's code, user manual, and demo: https://www.millisecond.com/download/library/mentalrotation/spatialprocessingtask/

i. Participants recruited from the Bard Community or otherwise through social media measures for the pilot study will receive $6.25 (for a 30-minute task and a New York State minimum wage of $12.50/hour as of January 1, 2021).
ii. Participants recruited from MTurk for the pilot study will receive $3.75 (for a 30-minute task at the U.S. Federal minimum wage of $7.25/hour).
iii. Participants recruited from MTurk for the screening questionnaire will receive $0.24 (for a 2-minute task at the U.S. Federal minimum wage of $7.25/hour).
iv. Participants recruited from MTurk for the main experiment will receive $2.50 (for a 20-minute task at the U.S. Federal minimum wage of $7.25/hour).

10. **Describe how you plan to mitigate (if possible) any risks the participants may encounter.** Since listening to audio and using electronic devices is a frequent occurrence for most participants, it is likely that they will be familiar with their comfort level during such tasks. Since listening to audio and using electronic devices is a frequent occurrence for most participants, it is likely that they will be familiar with their comfort level during such tasks. In order to further defray any potential discomfort, participants will:
    a. Be told in advance about the occurrence and length of exposure to these visual and auditory stimuli;
    b. Be able to set the volume to a comfortable level during the screening procedure, and
    c. Be offered short, self-directed breaks at various points of the procedures.
11. **Describe the consent process (i.e., how you will explain the consent form and the consent process to your participants):** All participants will be presented with the relevant consent form prior to the delivery of the task through Inquisit. The form contains general information about the goals of the experiment, risks and benefits, compensation, and confidentiality details, as well as the contact information for the primary investigator and the Bard Institutional Review Board.
12. **Have you prepared a consent form(s) and emailed it as an attachment to IRB@bard.edu?** Yes.
13. **If you are collecting data via media capture (video, audio, photos), have you included a section requesting consent for this procedure(s) in your consent form(s)?** N/A.
14. **If your project will require you to employ a verbal consent process (no written consent forms), please describe why this process is necessary and how verbal consent will be obtained and stored.** N/A.
15. **What procedures will you use to ensure that the information your participants provide will remain confidential and safeguarded against improper access or dissemination?** Participants will be assured during the consent and debriefing processes that steps are being taken to ensure the safeguarding of any personal information. Information collected through the Inquisit Web platform (answers to questionnaires and data from experimental tasks) will not be connected to their identifying information (for in-person participants, the name on their consent form; for MTurk participants, any information associated with their MTurk account) after the conclusion of data collection. All information provided through the Inquisit Web platform will be temporarily stored on their servers in Oregon, USA. Their security statement is accessible here. While it is stored there, it is only accessible to the primary investigator through their password-protected Inquisit account. At the conclusion of data collection, data will be stored offline on a password-protected computer to which only the primary investigator has access. Potentially

Pitch Interval as Context (2020NOV15-PAR, Amendment 2) 9

personal information collected (e.g., responses to the screening questionnaire) will only be used to determine qualification for further participation, and not connected to their scores in the main experiment, and all data (including demographics and other questionnaires, and task performance) will be reported in aggregate.

16. **Will it be necessary to use deception with your participants at any time during this research? Withholding details about the specifics of one's hypothesis does not constitute deception, this is called incomplete disclosure. Deception involves purposefully misleading participants about the nature of the research question or about the nature of the task they will be completing.** No.
17. **If your project study includes deception, please describe here the process you will use, why the deception is necessary, and a full description of your debriefing procedures.** N/A.
18. **For all projects, please include your debriefing statement. (This is information you provide to the participant at the end of your study to explain your research question more fully than you may have been able to do at the beginning of the study.) All studies must include a debriefing statement. Be sure to give participants the opportunity to ask any additional questions they may have about the study.** See Appendix C. for the debriefing statements.
19. **If you will be conducting interviews in a language other than English, will you conduct all of the interviews yourself, or will you have the assistance of a translator? If you will be using the assistance of a translator, that individual must also certify that he or she is familiar with the human subject protocol and has completed the online training course.** N/A.
20. **If your recruitment materials or consent forms will be presented in languages other than English, please translate these documents and email copies to IRB@bard.edu.** N/A.

Entry C6. IRB Approval, March 8, 2021.

Bard College Institutional Review Board

Date: March 8, 2021
To: Hadley Parum
Cc: Justin Hulbert, Deborah Treadway, Brandt Burgess
From: Tom Hutcheon, IRB Chair
Re: Proposed Amendments to 2020NOV15-PAR

DECISION: APPROVED

Dear Hadley,

The Bard Institutional Review Board has reviewed and approved the amendments you submitted to your protocol on March 8, 2021. Your case number remains 2020NOV15-PAR.

Please notify the IRB if your methodology changes or unexpected events arise.

This sounds like a really interesting project and we wish you the best of luck with your research!

Tom Hutcheon
IRB Chair
thutcheo@bard.edu

Appendix D. Preregistrations

Entry D1. Preregistration for Pilot I, submitted December 23, 2020.

Pilot Experiment [SUBMITTED 12/23/20]

1. **Data Collection. Have any data been collected for this study already?**
   a. Yes, we already collected the data.
   b. No, no data have been collected for this study yet.
   c. It's complicated. We have already collected some data but explain in Question 8 why readers may consider this a valid pre-registration nevertheless.
2. **Hypothesis. What's the main question being asked or hypothesis being tested in this study?** *Example: A month-long academic summer program for disadvantaged kids will reduce the drop in academic performance that occurs during the summer.*
   a. This study is being conducted in order to evaluate how changing the pitch interval of a polyrhythm affects participants performance in a task of semantic memory and their explicit subjective impressions of the polyrhythm. Investigating how pitch interval affects these responses will allow us to select four interval classes that are roughly equivalent in these respects, in order to include those particular sounds in a future study investigating context-dependent memory and pitch interval variance.
3. **Dependent variable. Describe the key dependent variable(s) specifying how they will be measured.** *Example: Simple average GPA across all courses during the first semester after the intervention.*
   a. There will be eight dependent variables collected, and two computed dependent variables to be used in analyses. For each 5:4 polyrhythm at its given pitch interval, participants will complete a lexical decision task (providing accuracy and reaction time) and provide explicit ratings about six extramusical features (whether a sound was familiar, pleasurable, consonant, engaging, distracting, and happy). For each participant at each pitch interval, a z-score will be computed for each of these variables with respect to the participant's overall mean on that variable.
   b. The global measure of abnormality will be a root sum squared (square root of the sum of squares of) the z-scores for each of the eight variables discussed above.
   c. The targeted measure of abnormality will be a root sum squared of a subset of the eight measured variables. A regression analysis will determine which of the six extramusical features were significant predictors of participant accuracy at each pitch interval. Only those significantly predictive musical features and LDT accuracy will be used to compute this rss metric.
4. **Conditions. How many and which conditions will participants be assigned to?** *Example 1: Two conditions: Offering summer program: yes vs no. Example 2: 12 conditions in a mixed design lab study. Participants will be assigned to one of four conditions: math training, verbal training, memory task, or control (4 between-subject conditions). Each participant will complete a math test, a verbal test, and a memory test (3 within-subject conditions).*
   a. The only factor determining conditions is the pitch interval at which participants hear the 5:4 polyrhythm. There are 36 potential pitch intervals, spanning all chromatic intervals from a minor second to three octaves apart. Each participant will hear 12 intervals throughout the experiment, such that they hear each possible interval class (e.g, a major third, or a perfect fifth) at one of three possible octave levels (i.e., one participant will hear notes that are two octaves and a major third apart, while another will hear notes that are a major third apart).
5. **Analyses. Specify exactly which analyses you will conduct to examine the main question/hypothesis.** *Example. Linear regression predicting the simple average GPA in the semester after the intervention with a dummy variable indicating whether the participant was offered the summer program or not (intention-to-treat analysis). We will also conduct the same regression controlling for simple average GPA during the semester before the intervention, gender, & household income (an 8-point scale ranging from 1 = below $20,000 and 8 = above $150,000).*
   a. The main analyses will consist of a 3x12 factorial ANOVA looking at the effect of octave level and interval class on the sound's relative abnormality, measured primarily by the **global** measure of abnormality, an rss metric. If evaluating multiple sets of four intervals that meet the above criteria, the targeted measure of abnormality will be used in order to break ties on any criteria.
   b. The selection of four interval classes will be guided by the following criteria:
      i. Included interval classes should not vary in their **targeted** abnormality according to octave level.
      ii. Either the perfect fourth or perfect fifth should be included, as this will be the pitch interval heard during the learning phase of the main experiment.
         1. If all interval classes or both the perfect fourth and fifth vary by octave level, interval classes should be selected that vary similarly by octave level (e.g., if the perfect fifth has a higher rss score at higher octave levels, no interval class should be selected that has a lower rss score at higher octave levels).
      iii. No two interval classes shall be selected that are musically inversions of each other (e.g., the major second and minor seventh are reciprocal intervals, so they would not both be included).
      iv. Ideally, one of the three additional interval classes would have a "consonant" relationship to either the 5/4 ratio of the polyrhythm or the 4/3 or 3/2 ratio (whichever of the perfect fourth or fifth have been selected), and two would have comparatively "dissonant" relationships to these ratios. It should not be the case that all three are consonant, or all three are dissonant.
      v. If possible, two of the three additional interval classes would be of equal musical distance from either the perfect fourth or fifth, whichever is selected (e.g., a minor third and a major seventh are both four semitones away from the perfect fifth, and would make for nice selections).
         1. Ideally, the dyad that is equidistant from this learning phase interval consists of one consonant and one dissonant interval, per item iv.
6. **Outliers and Exclusions. Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.** *Example 1. We will compute the overall mean and standard deviation across all conditions, and winsorize at 2.5 SD above/below the mean. Example 2: We will exclude participants who incorrectly answer at least 2 of our 3 attention check questions. Example 3. We will exclude any participants who complete the survey in less than 30 seconds.*
   a. Any incomplete responses, where participants terminated the program before completing the task, will be excluded.
   b. At the end of the pilot, participants will be asked whether they left audio playback audible throughout the study. All participants who report turning audio off at any point will not have their data included in analyses.
   c. Additionally, their data will be excluded based on responses to a short interval recognition task, where they are asked to r eport the direction (whether the second note is higher or lower than the first) and quality (e.g., a minor second or major sixth) of six intervals. If they are incorrect about the direction of four or more of these intervals, their data will be excluded. Additionally, if they correctly assess the quality of all six intervals, their data will be excluded.
7. **Sample Size. How many observations will be collected or what will determine sample size? No need to justify the decision, but be precise about exactly how the number will be determined.** *Example: We will offer the program until 500 people have agreed to participate in it or until June 30, 2016 (whichever comes first).*
   a. Data collection will continue until 72 participants have successfully completed the task or until February 15, 2021, whichever comes first.
8. **Other. Anything else you would like to pre-register?** (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?) *Example: We will include a battery of questions for exploratory purposes, including life satisfaction, amount of videogame playing, and family activity. We will also provide an additional survey with 24 questions assessing achievement orientation. We will not report the results of those analyses for the project being pre-registered. NOTE: If you leave this blank it will read 'Nothing else to pre-register.'*
   a. I will also be collecting information about participants' musical experience, and may use percentiles of musical experience (looking at years of formal training) to exclude musicians should they form a plurality of responses, in order to make selections most motivated by lay listeners.
   b. For reporting purposes, I will also be computing the numerical mean of the ratings of extramusical features, in order to be aware of how familiar, consonant, etc. participants were reporting the selected sounds to be.
9. **Name. Give a title for this AsPredicted pre-registration.** Suggestion: use the name of the project, followed by study description. Example: SUMMER PROGRAMS - GPA performance, Chicago, July 2018
   a. Varying Pitch Intervals: Parum Senior Project Pilot
10. **Finally. For record keeping purposes, please tell us the type of study you are pre-registering.**
   a. Class project or assignment
   b. Experiment
   c. Survey
   d. Observational/archival study
   e. Other:

Entry D2. Preregistration for Pilot II, submitted February 26, 2021.

Pilot Two [submitted 2/26/2021]

1. **Data Collection. Have any data been collected for this study already?**
    a. Yes, we already collected the data.
    b. No, no data have been collected for this study yet.
    c. **It's complicated. We have already collected some data but explain in Question 8 why readers may consider this a valid pre-registration nevertheless.**
2. **Hypothesis. What's the main question being asked or hypothesis being tested in this study?** *Example: A month-long academic summer program for disadvantaged kids will reduce the drop in academic performance that occurs during the summer.*
    a. The previous study "Varying Pitch Intervals: Parum Senior Project Pilot" found evidence supporting the selection of the minor third, perfect fourth, minor seventh, and perfect octave for a future experiment. However, the study also found irregularities in the ratings of distraction for all included sounds, with raw ratings hitting ceiling and z-scored ratings frequently reaching extreme values due to low variance in participants' responses. Accordingly, the present study will test whether refined musical stimuli, with a more naturalistic timbre, is able to counteract the levels of distraction and unpleasantness experienced by participants and support the selection of four interval classes for inclusion in a future experiment.
    b. This study is being conducted in order to evaluate how changing the pitch interval of a polyrhythm affects participants performance in a task of semantic memory and their explicit subjective impressions of the polyrhythm. Investigating how pitch interval affects these responses will allow us to select four interval classes that are roughly equivalent in these respects, in order to include those particular sounds in a future study investigating context-dependent memory and pitch interval variance.
3. **Dependent variable. Describe the key dependent variable(s) specifying how they will be measured.** *Example: Simple average GPA across all courses during the first semester after the intervention.*
    a. There will be eight dependent variables collected, and two computed dependent variables to be used in analyses. For each 5:4 polyrhythm at its given pitch interval, participants will complete a lexical decision task (providing accuracy and reaction time) and provide explicit ratings about six extramusical features (whether a sound was familiar, pleasurable, consonant, engaging, distracting, and happy). For each participant at each pitch interval, a z-score will be computed for each of these variables with respect to the participant's overall mean on that variable.
    b. The global measure of abnormality will be a root sum squared (square root of the sum of squares of) the z-scores for each of the eight variables discussed above.
    c. The targeted measure of abnormality will be a root sum squared of a subset of the eight measured variables. Because the previous pilot found that the musical feature ratings, in addition to reaction time and accuracy, were not normally distributed, I will not attempt to calculate a measure of targeted abnormality (this would have relied on a regression to determine which musical feature ratings significantly predicted differences in accuracy).
4. **Conditions. How many and which conditions will participants be assigned to?** *Example 1: Two conditions: Offering summer program: yes vs no. Example 2: 12 conditions in a mixed design lab study. Participants will be assigned to one of four conditions: math training, verbal training, memory task, or control (4 between-subject conditions). Each participant will complete a math test, a verbal test, and a memory test (3 within-subject conditions).*
    a. The only factor determining conditions is the pitch interval at which participants hear the 5:4 polyrhythm. There are 36 potential pitch intervals, spanning all chromatic intervals from a minor second to three octaves apart. Each participant will hear 12 intervals throughout the experiment, such that they hear each possible interval class (e.g., a major third, or a perfect fifth) at one of three possible octave levels (i.e., one participant will hear notes that are two octaves and a major third apart, while another will hear notes that are a major third apart).
5. **Analyses. Specify exactly which analyses you will conduct to examine the main question/hypothesis.** *Example. Linear regression predicting the simple average GPA in the semester after the intervention with a dummy variable indicating whether the participant was offered the summer program or not (intention-to-treat analysis). We will also conduct the same regression controlling for simple average GPA during the semester before the intervention, gender, & household income (an 8-point scale ranging from 1 = below $20,000 and 8 = above $150,000).*
    a. Because the previous iteration of this study found opted to use non-parametric tests in response to non-normal data and a small sample size, these data will similarly be analyzed using Kruskal-Wallis omnibus tests with Dunn's tests used for post hoc comparisons where significant differences are detected.
    b. The main analyses will investigate:
        i. Whether interval classes vary in overall abnormality according to octave level;
        ii. Whether in general, overall abnormality varied according to interval class or octave level;
        iii. Whether the normalized ratings the six musical features varied according to interval class or octave level.
    c. The procedure for selecting sounds will be similar to the previous study, with the following criteria:
        i. Included interval classes should not vary in overall abnormality according to octave level;
        ii. Either the perfect fourth or perfect fifth should be included;
        iii. Included interval classes should not be significantly different from each other in overall abnormality or according to any of the six musical features;
        iv. No two included interval classes should be musical inversions of each other (e.g., the major second and minor seventh are reciprocal intervals, so they would not both be included);
        v. Ideally, included interval classes include a balance of consonant and dissonant intervals;
        vi. Ideally, included interval classes include a balance of distances between the intervals.
6. **Outliers and Exclusions. Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.** *Example 1. We will compute the overall mean and standard deviation across all conditions, and winsorize at 2.5 SD above/below the mean. Example 2: We will exclude participants who incorrectly answer at least 2 of our 3 attention check questions. Example 3. We will exclude any participants who complete the survey in less than 30 seconds.*
    a. Any incomplete responses, where participants terminated the program before completing the task, will be excluded.
    b. At the end of the pilot, participants will be asked whether they left audio playback audible throughout the study. All participants who report turning audio off at any point will not have their data included in analyses.
    c. Additionally, their data will be excluded based on responses to a short interval recognition task, where they are asked to report the direction (whether the second note is higher or lower than the first) and quality (e.g., a minor second or major sixth) of six intervals. If they are incorrect about the direction of four or more of these intervals, their data will be excluded. Additionally, if they correctly assess the quality of all six intervals, their data will be excluded.
7. **Sample Size. How many observations will be collected or what will determine sample size? No need to justify the decision, but be precise about exactly how the number will be determined.** *Example: We will offer the program until 500 people have agreed to participate in it or until June 30, 2016 (whichever comes first).*
    a. Data collection will continue until 45 complete responses are recorded.
8. **Other. Anything else you would like to pre-register?** (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?) *Example: We will include a battery of questions for exploratory purposes, including life satisfaction, amount of videogame playing, and family activity. We will also provide an additional survey with 24 questions assessing achievement orientation. We will not report the results of those analyses for the project being pre-registered. NOTE: If you leave this blank it will read 'Nothing else to pre-register.'*
    a. With regards to the answer to question one indicating some data already being collected: This study is in response to the previous study ("Varying Pitch Intervals: Parum Senior Project Pilot"), and I do intend to compare results of both studies in order to see whether the timbre of musical stimuli used (a) has an effect on raw scores of musical features, especially distraction, and (b) the selection of sounds to be included in a future experiment. Question (a) will be analyzed using regression analyses, and (b) will be analyzed by performing the main analyses described in question 5. To that end, data has already been collected and analyzed for the previous study, which is the same in all aspects of procedure except for the timbre of the musical stimuli. No data has been collected yet using the updated musical stimuli.
    a. I also hope to perform exploratory analyses on both sets of data using regression analyses, in order to see whether ratings of musical features are predictable by altering the volume of audio, aspects of musical training, or the interval heard immediately prior to the rated interval.
9. **Name. Give a title for this AsPredicted pre-registration.** Suggestion: use the name of the project, followed by study description. Example: SUMMER PROGRAMS - GPA performance, Chicago, July 2018
    a. Varying Pitch Intervals: Parum Senior Project Pilot II
10. **Finally. For record keeping purposes, please tell us the type of study you are pre-registering.**
    a. Class project or assignment
    b. **Experiment**
    c. Survey
    d. Observational/archival study

Entry D3. Preregistration for Experiment, submitted March 10, 2021.

Registration Metadata

1. **Title.** Varying Pitch Intervals & Context-Dependent Memory: Parum Senior Project Experiment I
2. **Description.** Experimental portion of Hadley Parum's Senior Project, evaluating the effect of features of the pitch interval (interval class and octave level) of 5:4 polyrhythms on an expected context-dependent memory benefit to studied words.
3. **Contributors.** Hadley Parum.
4. **Category.** Experiment.
5. **Affiliated Institutions.** (You have no institutional affiliations.)
6. **License.** Creative commons (whatever you did for the other one)
7. **Subjects.** (whatever you did for the other one)
8. **Tags.** (whatever you did for the other one)

Study Information

1. **Hypothesis.** List specific, concise, and testable hypotheses. Please state if the hypotheses are directional or non-directional. If directional, state the direction. A predicted effect is also appropriate here. If a specific interaction or moderation is important to your research, you can list that as a separate hypothesis.
   a. The overarching goal of this research is to inquire whether manipulating the pitch interval of a 5:4 polyrhythm is sufficient to affect the expected context-dependent memory benefit.
   b. Specifically, the proportion of learned words recalled by participants in a delayed test is predicted to be higher for those (a) who hear the polyrhythm at the same pitch interval at test compared to those who hear no sound; (b) who hear the polyrhythm at the same octave level as that at learning [the first octave level], compared to those who hear more distant octave levels; (c) who hear the same interval class as that at learning [the perfect fifth], compared to different interval classes.
   c. Additionally, I hypothesize that musical experience may moderate this effect, with the effect of interval class being stronger for those with musical experience than without (i.e., those with musical experience may be more likely to experience the perfect fifth at octave level two or three as the "same" as the original sound).
   d. Finally, I hypothesize that explicit awareness of whether participants heard the same sound at both the learning and test timepoints may partially mediate this effect, such that whether a participant explicitly reported the sounds as the same is likely to better explain the effect octave level than interval class.

Design Plan

1. Study Type. Please check one of the following statements.
   a. **Experiment - A researcher randomly assigns treatments to study subjects, this includes field or lab experiments. This is also known as an intervention experiment and includes randomized controlled trials.**
   b. Observational Study - Data is collected from study subjects that are not randomly assigned to a treatment. This includes surveys, "natural experiments," and regression discontinuity designs.
   c. Meta-Analysis - A systematic review of published studies.
   d. Other
2. Blinding. Blinding describes who is aware of the experimental manipulations within a study. Mark all that apply.
   a. No blinding is involved in this study.
   b. **For studies that involve human subjects, they will not know the treatment group to which they have been assigned.**
   c. Personnel who interact directly with the study subjects (either human or non-human subjects) will not be aware of the assigned treatments. (Commonly known as "double blind")
   d. Personnel who analyze the data collected from the study are not aware of the treatment applied to any given group.
3. Is there any additional blinding in this study?
   a. No.
4. Study design. Describe your study design. The key is to be as detailed as is necessary given the specific parameters of the design. There may be some overlap between this question and the following questions. That is OK, as long as sufficient detail is given in one of the areas to provide all of the requested information. Examples include two-group, factorial, randomized block, and repeated measures. Is it a between (unpaired), within-subject (paired), or mixed design? Describe any counterbalancing required. *You may attach up to 5 file(s) to this question.*
   a. There is a between-subjects factorial design, where each participant will be randomly assigned to hear one of thirteen possible audio conditions at test, such that they either hear no sound or hear a 5:4 polyrhythm at one of 12 pitch intervals, classified by how they vary in interval class (major second, perfect fifth, major sixth, perfect octave) and octave level (within unison and one octave, one and two octaves, or two and three octaves).
5. Randomization. If you are doing a randomized study, state how you will randomize, and at what level. Typical randomization techniques include: simple, block, stratified, and adaptive covariate randomization. If randomization is required for the study, the method should be specified here, not simply the source of random numbers.
   a. Participants will be completing the list learning task in the Inquisit 6 Web player, with the background stimuli for the delayed test of word recall being randomly chosen from 13 files, including a silent file and loops of a 5:4 polyrhythm at 12 different pitch intervals. Condition information for each participant is randomly generated by Inquisit, without quotas from the experimenter. (This was chosen over having test sound condition be according to the participant's group number. Group numbers are generated when a potential participant loads the study's launch page in their browser. Previous piloting work revealed high rates of drop-out after loading the launch page, before starting the task. So, randomization based on group numbers may be uneven based on these drop-out rates.)

Sampling Plan

1. Existing Data. Preregistration is designed to make clear the distinction between confirmatory tests, specified prior to seeing the data, and exploratory analyses conducted after observing the data. Therefore, creating a research plan in which existing data will be used presents unique challenges. Please select the description that best describes your situation. See https://cos.io/prereg for more information.
   a. **Registration prior to creation of data**
   b. Registration prior to any human observation of the data
   c. Registration prior to accessing the data
   d. Registration prior to analysis of the data
   e. Registration following analysis of the data
2. Explanation of existing data. If you indicate that you will be using some data that already exist in this study, please describe the steps you have taken to assure that you are unaware of any patterns or summary statistics in the data. This may include an explanation of how access to the data has been limited, who has observed the data, or how you have avoided observing any analysis of the specific data you will use in your study.
   a. N/a.
3. Data collection procedures. Please describe the process by which you will collect your data and your inclusion and exclusion criteria. If you are using human subjects, this should include the population from which you obtain subjects, recruitment efforts, payment for participation, how subjects will be selected for eligibility from the initial pool, and your study timeline. For studies that don't include human subjects, include information about how you will collect samples, duration of data gathering efforts, source or location of samples, or batch numbers you will use. *You may attach up to 5 file(s) to this question.*
   a. All participants will be solicited from Amazon's Mechanical Turk (MTurk) platform, selecting from workers located in the U.S. with at least 90% approval rating for previous HITs. In order to access a pool of participants who (a) has not been diagnosed with any form of hearing loss, (b) does not consider themselves to have either of perfect pitch or tone deafness, (c) considers English a primary language, and (d) is willing and able to perform tasks in the downloadable Inquisit 6 Web platform, a separate screening HIT will be run on MTurk. The HIT will ask IRB-approved questions concerning eligibility for future research on various subjects, and participants deemed eligible will be re-contacted for the experimental HIT through encrypted MTurk IDs collected through CloudResearch's TurkPrime dashboard. All who participate in the screening HIT and provide a valid completion code will receive $0.25USD for a 2-minute task, and all who participate in the experimental HIT and provide a valid completion code will receive $2.50USD for a 15-20 minute task.
4. Sample size. Describe the sample size of your study. How many units will be analyzed in the study? This could be the number of people, birds, classrooms, plots, or countries included. If the units are not individuals, then describe the size requirements for each unit. If you are using a clustered or multilevel design, describe how many units are you collecting at each level of the analysis. This might be the number of samples or a range, minimum, or maximum.
   a. I plan to disburse the screening HIT to 780 unique participants, and aim to retain absolutely no fewer than 390 and no more than 559 participants through the experimental HIT. If within two weeks of inviting eligible participants to complete the experimental HIT fewer than 390 have done so, additional workers may be solicited through the screening HIT until at least 390 participants have completed the experimental HIT.
5. Sample size rationale. This could include a power analysis or an arbitrary constraint such as time, money, or personnel.
   a. The sample size will include a minimum of 390 participants and a maximum of 559 participants in the experimental portion of this study. Financial support is such that, with 780 participants recruited for the screening HIT, I can afford to compensate a maximum of 559 participants for their validated completion of the experimental HIT, after rounding to the nearest multiple of 13. A priori power analyses were conducted in G*Power for the two ANOVA tests and the regression test described in the analysis plan. For all three, a Bonferroni correction adjusted the alpha to 0.05/3 = 0.01666. The test with the largest sample size required to achieve a power level of .8 was the 4x3 factorial ANOVA, powering the factor of interval class (which has 4 levels) and estimating an effect size of .20. For this, 357 participants would be necessary in these 12 cells; so, 386.75 people would be necessary in the 13 experimental conditions in order to adequately power this ANOVA. This was rounded to the nearest higher multiple of 13, setting 390 as the minimum number of participants to be included in this experiment.
6. Stopping rule. If your data collection procedures do not give you full control over your exact sample size, specify how you will decide when to terminate your data collection. If you are using sequential analysis, include your pre-specified thresholds.
   a. Data collection will stop either when 559 participants have completed the experimental HIT and provided a valid completion code for payment, or on April 1.

Variables

1. Manipulated variables. Precisely define all variables you plan to manipulate and the levels or treatment arms of each variable. This is not applicable to any observational study. *You may attach up to 5 file(s) to this question.*
    a. The only manipulated variable will be the sound heard by participants during the delayed test of recall. They will either hear no sound, or hear a 5:4 polyrhythm at one of 12 pitch intervals. Additionally, these 12 pitch intervals will be classified according to their octave level and interval class.
2. Measured variables. Precisely define each variable that you will measure. This will include outcome measures, as well as any measured predictors or covariates. *You may attach up to 5 file(s) to this question.*
    a. The outcome measure of interest is participant memory of learned words at the delayed test of recall. This will be measured by the proportion of words correctly recalled divided by the total number of words. Correct recall is dependent on correct spelling, and exploratory analyses may address this.
    b. After the completion of the delayed test of recall, participants will be asked whether they believed the sound they heard at the test was the same as that heard during the learning phase, serving as a measure of explicit recognition.
    c. Additional variables measured in order to serve as descriptives of the population and as potential covariates include:
        i. Demographic information, including age in years, gender identity, and race/ethnicity.
        ii. Musical experience information, including whether and for how long a participant had musical training, whether and how frequently they play an instrument, and their amount of weekly musical listening.
    d. Additional variables measured primarily for observation and exclusion measures include:
        i. Validation questions about whether participants altered audio playback, how they completed the target tasks, and similar questions used to exclude those who illegitimately completed the task and to receive feedback on the task.
        ii. Performance in an interval recognition task, where participants hear two notes played sequentially and then simultaneously, and must judge whether the second note was higher or lower than the first (direction) and answer what formal interval name defined the pair (quality).
        iii. Performance in the manikin spatial recognition task, which serves as a distraction task between learning and final test.
3. Indices. If applicable, please define how measures will be combined into an index (or even a mean) and what measures will be used. Include either a formula or a precise description of the method. If you are using a more complicated statistical method to combine measures (e.g. a factor analysis), please note that here but describe the exact method in the analysis plan section. *You may attach up to 5 file(s) to this question.*
    a. N/a.

Analysis Plan

1. Statistical models. What statistical model will you use to test each hypothesis? Please include the type of model (e.g. ANOVA, RMANOVA, MANOVA, multiple regression, SEM, etc) and the specification of the model. This includes each variable that will be included, all interactions, subgroup analyses, pairwise or complex contrasts, and any follow-up tests from omnibus tests. If you plan on using any positive controls, negative controls, or manipulation checks you may mention that here. Provide enough detail so that another person could run the same analysis with the information provided. Remember that in your final article any test not included here must be noted as exploratory and that you must report the results of all tests. *You may attach up to 5 file(s) to this question.*
    a. First, testing whether the proportion of words recalled is higher for those who at test hear the same polyrhythm and pitch interval as during learning (compared to those who heard no sound), a one-way ANOVA will be performed on proportion of words recalled at final test with three groups: those who heard the same pitch interval, who heard a different pitch interval, and those who heard no sound.
    b. Second, testing the hypotheses about the effects of octave level and interval class, features of pitch interval, a 3x4 factorial ANOVA will be performed on proportion of words recalled at final test with octave level (1, 2, or 3) and interval class (major second, perfect fifth, major sixth, and perfect octave) as the relevant factors; this analysis will exclude participants who heard no sound at test.
    c. Exploring the potential effects of moderating variables, regression analyses will be performed on the proportion of words recalled at the final test. Differences in R^2_{adj} will be computed between Model 1-2 and Model 2-3.
        i. Model one (Pitch Interval): Interval Class, Octave Level, and their interaction if it is significant in the factorial ANOVA.
        ii. Model two (Pitch Interval and Musical Experience): Model one predictors and whether participants reported experiencing musical training (yes/no) or currently playing music (yes/no).
        iii. Model three (Pitch Interval and Music Playback): Model one predictors and whether participants reported changing volume at any point during the task (yes/no) and how participants reported listening to audio during the task (headphones/speakers/other).
    d. Mediation analyses will assess the potential mediating factor of explicit recognition of the test sound on the relationship between context change and memory for words.
2. Transformations. If you plan on transforming, centering, recoding the data, or requiring a coding scheme for categorical variables, please describe that process.
    a. N/A.
3. Inference criteria. What criteria will you use to make inferences? Please describe the information you'll use (e.g. specify the p-values, Bayes factors, specific model fit indices), as well as cut-off criterion, where appropriate. Will you be using one or two tailed tests for each of your analyses? If you are comparing multiple conditions or testing multiple hypotheses, will you account for this?
    a. The present study will use null hypothesis significance testing with a p-value below .05 being considered significant.
4. Data exclusion. How will you determine which data points or samples if any to exclude from your analyses? How will outliers be handled? Will you use any awareness check?
    a. Participants' data will be removed from analysis of the experimental HIT if any of the following are the case:
        i. If Inquisit reported them as not completing the task, due to manually quitting the program or another technical difficulty resulting in incomplete data.
        ii. If their responses to the interval recognition test demonstrate behavior in line with inattention and/or tone deafness and/or perfect pitch, such that:
            1. They fail to correctly identify the direction of more than three of the six intervals; or if
            2. They successfully name the formal interval name of all six intervals.
        iii. If their responses to the validation questions presented at the end of the study are such that:
            1. They report expecting the delayed test of recall;
            2. They report keeping physical or typed notes of the studied words;
            3. They report turning off the audio playback at any point of the study.
        iv. If their responses to the manikin spatial recognition task demonstrate behavior in line with inattention, such that
            1. They complete all 680 trials before the 240 second timeout; or
            2. There was at least one period where they took longer than 30 seconds to make a judgement.
5. Missing data. How will you deal with incomplete or missing data?
    a. Responses will not be included in analyses if the participant did not fully complete the task, as indicated by Inquisit. All measures relevant to analyses force a participant response.
6. Exploratory analysis. If you plan to explore your data to look for unspecified differences or relationships, you may include those plans here. If you list an exploratory test here, you are not obligated to report its results. But if you do report it you are obligated to describe it as an exploratory result.
    a. The measure of final recall may be recoded to include misspelled words, in order to ensure that this doesn't significantly alter results.
    b. Since pilot experimentation generated information about the pleasantness, etc. of the sounds included in this study, and participants are asked to rate words for pleasantness, exploratory analyses may investigate whether participants tended to recall words whose pleasantness was similar to that of the test sound.

Other

If there is any additional information that you feel needs to be included in your preregistration, please enter it here. Literature cited, disclosures of any related work such as replications or work that uses the same data, or other helpful context would be appropriate here.