BIDIRECTIONAL INFLUENCES OF PITCH AND TIME

BIDIRECTIONAL INFLUENCES OF PITCH AND TIME IN AUDITORY PERCEPTION

By JESSE KENDALL PAZDERA, B.Sc.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy

August 23, 2024

McMaster University

DOCTOR OF PHILOSOPHY (2024)

Hamilton, Ontario, Canada (Psychology, Neuroscience & Behaviour)

TITLE:	Bidirectional Influences of Pitch and Time in Auditory Perception
AUTHOR:	Jesse Kendall Pazdera B.Sc. (Psychology & Computer Science), Haverford College, Haverford, PA, USA
SUPERVISOR:	Dr. Laurel J. Trainor
NUMBER OF PAGES:	xviii, 198

Lay Abstract

Our ability to understand rhythms and find "the beat" in music and speech is key to how we interact with the world and with one another. Rhythm and music are important in every known culture, and synchronizing to rhythms helps us form connections, coordinate, and communicate with others. This thesis explores how another aspect of music—pitch—changes how we hear the beat. Past research suggests music sounds faster to us when played at a higher pitch. Through our work, we discovered that the reverse is also true—musical pitch starts to sound higher as the rhythm speeds up. We also show that pitch changes how fast we move while trying to keep the beat. Studying these pitch and rhythm illusions helps us to better understand how our brains combine information about the melody and rhythm of music, and may help us to develop better medical alarms in the future.

Abstract

Auditory rhythms play a central role in human culture and communication, through both speech and music. The ability to track and predict the organization of events in time helps humans optimize attention, perceive emotion, coordinate actions, and understand social affiliations. The importance of these functions has inspired substantial efforts to model rhythm perception. However, despite a wealth of evidence that pitch influences rhythm perception, with higher speech and music perceived as faster, leading theories and models of rhythm perception have yet to incorporate these effects of pitch. This thesis addresses several empirical questions that have stood in the way of integrating pitch into these models. Specifically, 1) whether the perception of higher pitches as faster generalizes across more than two octaves and above 1000 Hz, 2) whether pitch influences synchronized motor tempo, and 3) whether pitch–timing interactions are bidirectional, such that tempo changes also influence perceived pitch. To answer these questions, we present data from ten experiments including subjective tempo ratings, sensorimotor timing, temporal discrimination, and pitch discrimination tasks. Our results suggest the existence of two separate effects of pitch on perceived timing. First, we present evidence in Chapters 2 and 3 for a unidirectional, negative quadratic effect of absolute pitch on perceived tempo. In this effect, both subjective and sensorimotor tempo rise with pitch between 110 and 440

Hz, peak somewhere between 440 and 1760 Hz, and decrease with pitch above that peak. In Chapters 4 and 5, we present evidence for a bidirectional and approximately linear bias to perceive higher pitches as faster and earlier sounds as higher. We propose that the former effect is most likely innate and a product of the structure of the auditory system, whereas the latter is learned from world structure and originates from cue integration at a later stage of processing.

To My Mother

I'm everything I am because you loved me

Acknowledgments

There are so many individuals without whom I would have never reached this point. My first year of graduate school was marked by the loss of my mother, followed one month later by the onset of the COVID-19 pandemic. The support of so many friends, family members, and mentors has allowed me to see this journey through to the end, in spite of all we lost.

Thank you, Laurel, for all these years of mentorship and support. Somehow you always found time to be there for me, despite having recruited an overwhelming four new PhD students in my first year. Your care for each and every one of us has not gone unnoticed, especially throughout the pandemic. I'm not sure my research would have made it through Winter 2020, if not for your support.

Thank you, as well, to Marilyn Boltz for introducing me to music cognition research ten years ago, as I was first realizing I didn't want to join the tech industry. You saw something in me, and helped me find my place when I was feeling lost. Ever since, you have been my mentor and my friend, and I hope that my continuation of your research on pitch and timing has made you proud. In a way, this whole thesis is an expression of my gratitude.

Thank you to the members of my PhD committee—Mike Schutz and Ed Large, and my former NSERC–CREATE advisor Caroline Palmer. Thank you also to Jon

Cannon for serving on my Comprehensive Exam committee, and for collaborating with Ed, Laurel, and I on quite a monster of a review paper. It has been an honor to work with and learn from such an amazing team of scientists throughout the past five years. And thank you to Mike Kahana for the guidance that allowed me to assemble such a wonderful team of mentors.

Thank you to Gabriel Xiao, Ali Hashemi, and Ayesha Khan for helping me to become a better teacher and mentor myself. I started graduate school thinking I wasn't cut out for teaching, and now it's the most rewarding part of the work I do. And to all of my students who have helped bring my research ideas to life: Kian Yousefi Kousha, Habiba Imam, and Kunika Singh, Colin Bruce, Maggie Wu, Natalie Jean-Marie, Sharanya Badalera, Emily Melo, Olive Rinaldi, Caitlyn Humber, and Ellia Baines. You have all made "Jesse's Lab" into something truly special.

Thank you to my mother and father for being the most loving parents I could have ever hoped to have. You helped me through so many anxieties, allergies, and doubts over my whole life, ever since I was a little child afraid of the lines on a grilled hot dog. You taught me to be sensitive, caring, and kind, and always encouraged me to do what I love. Following my heart has led me to where I am today, and you have been there every step of the way. Mommy, although you're no longer here in body, I know you'll be there for me always. I'm everything I am because you loved me.

Grandmom and PopPop, you gave me so many opportunities in life that I never would have had otherwise. Words could never express how much you meant to me, but I know that watching me grow and succeed was the only thanks you ever needed. Grandma Karen, I'm not sure I would have come back to Canada when Mommy passed away if you weren't there for my dad and I through the worst of times. And

to the original Dr. J. Pazdera: You were building computational models in the 1960s on a computer the size of a room. Though we never met, I like to think that some part of your scientific mind made its way to me. Through my work, I hope to keep your legacy alive.

Thank you to my dear friends Matt Beal and Effie Li (李俞萱) for being there for me in 2020 throughout the most difficult year of my life. And thank you to Sigal Balshine and Shira Miller for helping me to reconnect with my cultural community after losing my mother and grandparents. Thank you to my House Bubble friends Emily Wood, Konrad Swierczek, Lucas Klein, Leigh Greenberg, Elizabeth Phillips, and later Zoë Lazar–Kurz for all your company throughout the pandemic, and to Maggie for being the sweetest and most loving kitty I have ever met.

Thank you to all my other labmates for your friendship, feedback, guidance, and support throughout my years here: Elaine Whiskin, Susan Marsh–Rollo, Sally Stafford, Rachael Finnerty, Erica Flaten, Chantal Carrillo, David Prete, Shu Sakamoto, Sara Ripley, Maya Psaris, Dan Cameron, Sean McWeeny, Martín Miguel, Dobri Dotov, and Dan Bosnyak. And thank you to Dave Thompson for all your help building the tapping apparatus for Chapter 3.

Thank you to my other friends in the program and around the world. Hasan Siddiqui for tirelessly keeping the social spirit of our department alive. Ben Sclodnick, Shane Seheult, Rohit Singh, Brendan Ward, Brendan McEwen, and Rafa Román-Caballero for being my first MTG community in almost fifteen years. Esin Gürcan for so much good food, gossip, and laughs. Meg Lall and Selena Singh for many long walks, study sessions, and hearts-to-heart. Casey Falk and David Reilley for all our video game nights and vacations together. Jocelyn De La Cruz for endless emotional

support over the past year. And a special thank you to Teagan Lauriente, for spending days of your life combing through my ideas with me, until I could somehow organize them into the (generally) coherent document I present today. Some of my best ideas came out of our conversations, and I hope that I may give you the same. Each and every one of you allowed me to see this journey through to the end. Thank you.

Table of Contents

Lay Abstract **iii**

Abstract **iv**

Acknowledgments **vii**

Abbreviations and Symbols **xvii**

Declaration of Academic Achievement **xix**

1 General Introduction **1**

1.1 Rhythm as a Perceptual and Social Cue 2

1.2 Models of Rhythm Perception 3

1.3 Perceptual Integration of Pitch and Time 5

1.4 Origins of Pitch–Timing Interactions 10

1.5 Goals of the Present Work 14

2 Pitch-induced illusory percepts of time **17**

2.1 Preface 17

2.2 Introduction 18

2.3 Experiments 1–3 . . . 25
2.4 Experiment 4 . . . 47
2.5 Experiment 5 . . . 55
2.6 General Discussion . . . 62

3 Pitch biases sensorimotor synchronization to auditory rhythms 66

3.1 Preface . . . 66
3.2 Introduction . . . 67
3.3 Experiment 1 . . . 74
3.4 Experiment 2 . . . 88
3.5 General Discussion . . . 95
3.6 Conclusion . . . 101

4 Bidirectional interactions of pitch and time 102

4.1 Preface . . . 102
4.2 Introduction . . . 103
4.3 Experiment 1 . . . 105
4.4 Experiment 2 . . . 112
4.5 General Discussion . . . 116

5 Timing-induced illusory percepts of pitch 118

5.1 Preface . . . 118
5.2 Introduction . . . 119
5.3 Experiment 1 . . . 122
5.4 Experiment 2 . . . 135
5.5 General Discussion . . . 144

5.6 Conclusion 150

6 General Discussion **151**

6.1 Summary of Findings 153

6.2 Theoretical Implications and Open Questions 155

6.3 Implications for Models of Rhythm Perception 164

6.4 Practical Applications for Auditory Interfaces 171

6.5 Concluding Remarks 173

References **174**

List of Figures

2.1 Factors Varying Across Octaves 24

2.2 Example Data Processing 32

2.3 Raw Tempo Ratings in Experiments 1–3 39

2.4 Illusory Tempo by Pitch Height in Experiments 1–3 41

2.5 Illusory Tempo Effects by Tempo Range in Experiments 1–3 42

2.6 Sensorimotor Synchrony and the Illusory Tempo Effect in Experiments 1–3 43

2.7 Raw Tempo Ratings in Experiment 4 51

2.8 Illusory Tempo by Pitch and Timbre in Experiment 4 52

2.9 Raw Tempo Ratings in Experiment 5 58

2.10 Illusory Tempo Effects by Register in Experiment 5 60

3.1 Effect of Pitch on Subjective Tempo (Pazdera & Trainor, 2024b) . . . 70

3.2 Illustration of the Arduino's Pre-Stimulus "Blind Spot" 80

3.3 Example Asynchrony Calculation for One Participant 82

3.4 Effect of Pitch on Sensorimotor Timing 86

4.1 Pitch Biases Perceived Mistiming 108

4.2 Effect of Preceding Context on Perceived Mistiming. 110

4.3 Timing Biases Perceived Pitch 115

5.1 Probability of Rating Probes as Higher Than the Standard 127
5.2 Sensitivity and Bias of Pitch Discrimination 129
5.3 Individual Differences in d' and Timing-Induced Bias 130
5.4 Reaction Times for Correct Responses 131
5.5 Pitch Discriminability by Difficulty Level 140
5.6 Timing-Induced Bias by Difficulty Level 141
5.7 Individual Differences in JND and Timing-Induced Bias 142
6.1 Hypothetical Oscillation of Perceived Pitch During Entrainment . . . 168

List of Tables

1.1 Evidence for Interactions of Pitch and Time 13

6.1 Revised Evidence for Perceptual Interactions of Pitch and Time . . . 152

Abbreviations and Symbols

Abbreviations

ANOVA	Analysis of variance
BPM	Beats per minute
DV	Dependent variable
FSR	Force sensitive resistor
IOI	Interonset interval
IV	Independent variable
JND	Just-noticeable difference
MANOVA	Multivariate analysis of variance
MEG	Magnetoencephalography
NRT	Neural Resonance Theory
PATIPPET	Phase and Tempo Inference from Point Process Event Timing

Symbols & Notation

C	Measure of bias (signal detection theory)
d'	Sensitivity index (signal detection theory)
$f0$	Fundamental frequency
M	Mean
N	Number of participants
SD	Standard deviation
SE	Standard error of the mean

Declaration of Academic Achievement

I, Jesse Pazdera, declare that this thesis entitled "Bidirectional influences of pitch and time in auditory perception" and the work presented herein are my own. I am the sole author of the Introduction and General Discussion chapters. Chapters 2 through 4 were originally drafted exclusively by myself, with editing and consultation by Laurel J. Trainor. The text of Chapter 5 was originally drafted exclusively by myself, with data analysis and visualization support by Olive M. Rinaldi, and editing and consultation by Laurel J. Trainor. The tapping apparatus used in Chapter 3 was designed and constructed in consultation with Dave Thompson. The LaTeXtypesetting format of the present document is derived from the unofficial McMaster University thesis template developed by Omar Boursalie and Asif Khan: `https://www.overleaf.com/latex/templates/mcmaster-university-thesis-template/yzxqvhkmhgjk`.

The format of this thesis is as a sandwich thesis, and the works that comprise Chapters 2–5 have all been prepared as manuscripts for publication in academic journals. At the time of thesis submission, Chapters 2 and 3 have been submitted for publication, and are currently under review. Chapter 4 has been published in a conference proceedings. Chapter 5 will be submitted for publication later this year.

Chapter 1

General Introduction

Auditory rhythms play a central and universal role in culture and communication through both speech and music (S. Brown & Jordania, 2011; Jacoby et al., 2024; Mehr et al., 2019). Here, I use *rhythm* to refer to any pattern of events that is predictably organized over time, such as the syllables of speech or the notes of a song. These events are often modeled as discrete, but need not be so (Large et al., 2023); for example, the predictable fluctuations in intensity of a continuous amplitude modulated tone can be perceived as rhythmic (e.g., Herrmann & Johnsrude, 2018). Humans are able to extract and synchronize to a perceived *beat* or *pulse* of regularly-spaced events underlying auditory rhythms, as are some non-human animals (see Wilson & Cook, 2016 for a review), including certain species of monkeys (Gámez et al., 2018), apes (Large & Gray, 2015), cockatoos (Patel, Iversen, Bregman, & Schulz, 2009a), and sea lions (Cook, Rouse, Wilson, & Reichmuth, 2013). The frequency of this perceived pulse is referred to as the rhythm's *tempo*, and in humans, the ability to accurately track the tempo of auditory rhythms is critical for both perception and social cognition.

1.1 Rhythm as a Perceptual and Social Cue

Predictable temporal structures make it possible to anticipate when upcoming events will occur. Dynamic Attending Theory (Jones, 1976; Jones & Boltz, 1989; Large & Jones, 1999) suggests that this predictability enables the brain to enhance perception by focusing attention at points in time when critical information is most likely. Jones (1976) originally hypothesized that endogenous neural oscillations termed "perceptual rhythms" synchronize to the temporal patterns of external stimuli. Jones and Boltz (1989) and Large and Jones (1999) formalized the concept of perceptual rhythms into a model in which attention phase-locks or *entrains* to periodic stimuli in such a way as to optimize the distribution of perceptual resources over time. This optimization may explain why rhythmic priming can enhance performance on tasks such as phonetic (Cason & Schön, 2012) and semantic (Rothermich, Schmidt-Kassow, & Kotz, 2012) processing, pitch (Chang, Bosnyak, & Trainor, 2019) and duration (McAuley & Fromboluti, 2014) perception, and auditory gap detection (Henry & Herrmann, 2014; Henry, Herrmann, & Obleser, 2014).

In addition to facilitating perception, the ability to extract tempo from the rhythm of communication is an important tool for social cognition. Tempo is a key indicator of emotion in both speech (Scherer, 1986) and music (Scherer & Oshinsky, 1977). Specifically, faster timing tends to convey and induce emotions with more positive valence and greater intensity, compared to slower timing (Balkwill & Thompson, 1999; Battcock & Schutz, 2019; Collier & Hubbard, 1998, 2001; Gabrielsson & Juslin, 1996; Juslin, 1997; Liu et al., 2018; Trochidis & Bigand, 2013). Speaking rate also informs perceived personality, with faster speakers being perceived as less passive (Apple, Streeter, & Krauss, 1979) and more extraverted (Addington, 1968)—and indeed speaking rate has been demonstrated to be predictive of a person's extraversion (Mairesse & Walker, 2006). Faster speakers are also perceived as more confident (Guyer, Fabrigar, & Vaughan-Johnston, 2018), competent (B. L. Brown, Strong, & Rencher, 1973; Ray, 1986; Smith, Brown, Strong, & Rencher, 1975), and credible (Miller, Maruyama, Beaber, & Valone, 1976) than slow speakers, while at the same time faster speakers may be perceived as less benevolent (B. L. Brown et al., 1973; Ray, 1986).

Collective entrainment to a beat also has important social consequences. Cross-culturally, workers have long used song as a tool for synchronizing labor (Gioia, 2006; Korczynski, 2003; Poppe, 2010). This form of collective beat extraction and synchronization has also been shown to increase feelings of affiliation between synchronized individuals (Hove & Risen, 2009), and to increase subsequent group cooperation in economic games (Wiltermuth & Heath, 2009). More recent research suggests that infants and children use interpersonal synchrony as a cue to social affiliations between themselves and others (Cirelli, 2018; Cirelli, Einarson, & Trainor, 2014; Cirelli, Trehub, & Trainor, 2018), as well as between third-parties (Cirelli, Wan, Johanis, & Trainor, 2018). Therefore, being able to quickly and accurately assess the temporal structure of the environment conveys social and perceptual advantages from an early age.

1.2 Models of Rhythm Perception

Given the universal cultural importance of rhythm and synchrony, substantial efforts have been made to develop models of rhythm perception and understand the process by which humans (and some nonhuman animals) are able to extract the beat of auditory time series. Current models generally fall under one of two categories: predictive coding/Bayesian inference, or neural resonance (Large et al., 2023; Palmer & Demos, 2022).

Karl Friston's (2009, 2010) introduction of the free-energy principle as a unified theory of perception has inspired the development of predictive coding accounts of rhythm perception (Koelsch, Vuust, & Friston, 2019; Vuust, Dietz, Witek, & Kringelbach, 2018; Vuust & Witek, 2014). Predictive coding accounts consider the brain to be a Bayesian prediction engine that attempts to develop a generative model of its environment. It develops such a model by continuously generating top-down predictions of incoming sensory information, and then minimizing the error of its predictions by back-propagating the difference between its predictions and observations (Friston & Kiebel, 2009). In the context of rhythm perception, it is hypothesized that the brain attempts to infer the phase and tempo of rhythmic stimuli based on the onset times of events, as a means of predicting future events (Cannon, 2021).

In contrast to explicit prediction, Neural Resonance Theory instead proposes that rhythm perception is an emergent property of the physical dynamics of neural oscillators (Large, 2010; Large & Snyder, 2009). Neural networks in the sensory and motor cortices are hypothesized to resonate when rhythmically stimulated, producing oscillatory activity at frequencies present within the stimulus, as well as at harmonics and subharmonics of those frequencies (Large, Almonte, & Velasco, 2010). These networks are structured as gradient frequency networks, consisting of arrays of neural oscillators that are each tuned to different preferred tempos. The perceived pulse and tempo of auditory rhythms depends on the frequencies of the oscillators that resonate most strongly in the network—which, due to the nature of complex dynamics, may or may not be frequencies that were present in the original stimulus (Large, Herrera, & Velasco, 2015; Tal et al., 2017). Through Hebbian learning, these gradient frequency networks become attuned to the rhythmic patterns most commonly found in one's environment and native musical culture (Tichko, Kim, & Large, 2022; Tichko & Large, 2019).

Currently, neither of these theories of rhythm perception explicitly incorporates information about the pitch of an auditory time series when extracting the beat. However, it has long been hypothesized that pitch and time are integrated in auditory perception (Cohen, Hansel, & Sylvester, 1954; Jones, 1976; Boltz, 1999), and evidence for this integration remains under-addressed in our theories and models of rhythm perception.[1]

1.3 Perceptual Integration of Pitch and Time

Pitch refers to the subjective quality of how high or low a sound is, and can be approximately defined as the perceived fundamental frequency of the sound (which, similar to tempo perception, may or may not be a frequency present in the sound's spectrum; Bendor & Wang, 2005). Whereas pitch perception and rhythm perception have largely been studied independently in music cognition research, Jones (1976) proposed that the brain integrates the pitch and timing of auditory signals into a single abstract representation, rather than processing each independently. She hypothesized that the brain represents music as a trajectory through this combined space, and that

[1] Two exceptions may be noted: Cannon (2021) proposed extending his Bayesian model to generate more precise phase inferences for lower voice lines in polyphonic music, and Large (2000) conducted a simulation in which lower pitches coupled more strongly to slower oscillators in a gradient frequency network.

this trajectory can be used to make predictions about how the music will continue to unfold in pitch and time. The structure of this abstract space reflects the structure of sounds in the real world, such that lawful relations between pitch and time in nature also constrain expectations of movement through the space. In particular, Jones (1976) focused on the proportionality of change as one such lawful relation. Larger changes in pitch tend to occur over larger periods of time; therefore, a change in pitch between two sounds carries some information about how much time elapsed between them, while the elapsed time between two sounds carries some information about how far apart they were in pitch (Cohen et al., 1954; Henry & McAuley, 2009, 2013). In addition to proportionality, it has been suggested that absolute pitch and/or the directionality of pitch changes may be correlated with absolute tempo and/or the directionality of tempo change (e.g., Boltz, 1998, 2011, 2017; Broze & Huron, 2013; Feldstein & Bond, 1981; Gordon & Ataucusi, 2021), and it is on these relationships that the present empirical work will focus.

To the extent that movement in pitch is informative about movement in time, the brain should—from an optimal Bayesian perspective (Friedman, Ludvig, Legge, & Vuong, 2013)—incorporate pitch as a cue to the phase and tempo of musical and vocal rhythms. Although Jones's original framework did not use the language of predictive coding, both theories center on the brain constructing an internal model that reflects the statistical structure of the outside world. Relatedly, in Neural Resonance Theory, statistical relations between pitch and tempo might also be internalized through Hebbian learning, as part of the process of attunement to one's environment (Tichko et al., 2022). The integration of relevant non-temporal cues into rhythm perception might therefore support its important role in communication and social cognition.

Regardless of the mechanism, if pitch and time are perceptually integrated and real-world correlations exist between them, then we should expect to find that pitch influences the perceived timing of sounds, and vice versa. Indeed, substantial evidence exists to suggest that both absolute pitch and pitch change can produce illusory changes in perceived tempo, although the inverse effects of timing on perceived pitch have received relatively little investigation. Table 1.1 summarizes the existing evidence for bidirectional interactions between pitch and tempo in auditory perception, and prefaces the interactions that will be investigated further in the later chapters of this manuscript.

1.3.1 Effects of Pitch on Perceived Tempo

Higher absolute pitch has consistently been associated with faster timing or speeding up. Early work by Feldstein and Bond (1981) asked participants to rate which of two speech samples was faster, while pitch-shifting one of the samples by approximately two semitones. They found that participants tended to rate whichever speech sample was higher as being faster. Boltz (2011) later conducted a similar study using short melodies in place of speech, and found that people also perceived melodies as faster when played in a higher octave (starting on C5 = 523.3 Hz) than when played in a lower octave (starting on C3 = 130.8 Hz). Instead of a two-alternative forced choice task, Collier and Hubbard (1998) asked participants to rate on a Likert scale the speed and tempo change of repeating tones (C4 = 261.6 Hz, C5, or C6 = 1046.4 Hz) and C major scales (between C4 and C5 or C5 and C6). Their participants similarly rated higher-octave tones and scales as faster and speeding up more than those played in a lower octave. Most recently, Boltz (2017) conducted a voice recognition experiment in which participants were familiarized with a set of male and female voices, and later exposed to 10% pitch- and/or tempo-shifted variants of those voices. Participants were most accurate at recognizing that a voice's tempo was modified when its pitch was also modified in the congruent direction, suggesting an integrated memory coding of pitch and time.

In a somewhat different design testing the perceived duration of short ($\approx$ 200 ms) single intervals, Pfeuty and Peretz (2010) found that the pitch of both flanking tones affected perceived duration. Higher tones produced shorter perceived intervals, regardless of whether that tone initiated or concluded the interval. Pitches as little as one quarter semitone apart were sufficient to produce different perceived interval lengths, although individuals with amusia did not show a consistent pitch-induced timing bias, even between pitches as much as four semitones apart. Together, these results suggest that higher-pitched sounds are perceived as faster both within a single octave (Boltz, 2017; Feldstein & Bond, 1981; Pfeuty & Peretz, 2010) and across octaves (Boltz, 2011; Collier & Hubbard, 1998), and that these effects apply to both single intervals (Pfeuty & Peretz, 2010) and rhythmic timing (e.g., Collier & Hubbard, 1998), in both speech (Feldstein & Bond, 1981; Boltz, 2017) and music (Boltz, 2011).

In addition to absolute pitch, pitch changes have likewise been found to induce illusory tempo changes in the congruent direction. For example, the aforementioned studies by Boltz (2011) and Collier and Hubbard (1998) also observed that participants rated ascending scales and melodies as faster and speeding up more than descending ones. Similarly, Lake, LaBar, and Meck (2014) found that participants were biased to rate a repeating tone as slowing down if its pitch decreased on the final two repetitions, but speeding up if its pitch increased. Several studies have since demonstrated effects of continuous pitch glides on perceived tempo, as well. Herrmann and colleagues showed that applying a pitch glide to a frequency-modulated tone produced a perceived change in modulation rate, such that participants tended to perceive ascending pitch glides as speeding up and descending pitch glides as slowing down (Herrmann, Henry, Grigutsch, & Obleser, 2013; Herrmann, Henry, Scharinger, & Obleser, 2014). They later replicated this effect in amplitude-modulated tones (Herrmann & Johnsrude, 2018), while Gordon and Ataucusi (2021) replicated it with pitch glides applied to jazz melodies.

Initial evidence also suggests that the direction of pitch change affects both neural and sensorimotor synchronization to rhythmic sequences. One magnetoencephalography (MEG) study identified a difference in the relative phase, but not frequency, of stimulus-entrained oscillations in the auditory cortex, depending on whether the pitch of a frequency-modulated tone ascended or descended over time (Herrmann et al., 2013). Furthermore, intertrial phase coherence was greater when pitch and tempo changed in the same direction than when they changed in opposite directions, suggesting that congruent pitch changes improved neural tracking of tempo changes. As for sensorimotor synchronization, Boasson and Granot (2012) found in a synchronization tapping task that participants began tapping faster when pitch ascended, compared to when it descended. In combination, these results suggest that pitch–timing integration most likely occurs upstream of the motor system, manifesting in the neural dynamics of the auditory cortex. However, further study of both the neural and sensorimotor effects of pitch on rhythm perception is warranted, given that the vast majority of research has focused exclusively on subjective ratings of tempo.

1.3.2 Effects of Tempo on Perceived Pitch

In contrast to the effects of pitch on perceived tempo, the effects of timing on perceived pitch have not been well-studied. Most evidence comes from a single series of experiments by Madsen and colleagues (Duke, Geringer, & Madsen, 1988; Geringer & Madsen, 1984; Madsen, Duke, & Geringer, 1984). In each of three studies, they presented participants with two consecutive recordings of the same orchestral (Duke et al., 1988; Geringer & Madsen, 1984) or band (Madsen et al., 1984) performance, in which the tempo and pitch of the latter recording were shifted up or down. In general, participants rated the directionality of the tempo alteration more accurately when the pitch was also modified in the congruent direction (pitch increase and faster tempo, or pitch decrease and slower tempo) than when pitch and tempo were modified in opposite directions. Boltz (2017) later replicated this effect in speech excerpts. These results suggest that faster tempo is associated with higher perceived pitch, but further investigation is needed into whether tempo changes also produce illusory changes in pitch, as would be expected if pitch and time are perceptually integrated (Jones, 1976).

1.4 Origins of Pitch–Timing Interactions

Substantial evidence supports the notion that pitch influences tempo perception—and possibly vice versa—but before pitch can be incorporated into models of rhythm perception, a clearer understanding of the origins of the effect are needed. Broadly speaking, there is a general consensus among researchers that pitch–timing interactions arise from learned correlations that reflect lawful relations in real-world structure, as Jones (1976) hypothesized. However, there have been diverse suggestions regarding from where this correlation derives.

The most common hypothesis is that pitch and tempo covary in human communication due to changes in emotional arousal (Boltz, 2011; Collier & Hubbard, 1998; Feldstein & Bond, 1981; Lake et al., 2014). Both higher pitch and faster tempo are forms of higher-frequency oscillation, and may correlate with higher arousal. At least one experiment has shown that as people increase the loudness of their voice, they also tend to start speaking faster and in a higher pitch (Black, 1961). However, evidence that pitch is associated with arousal in music is mixed (Battcock & Schutz, 2019). To the extent that pitch and tempo do tend to covary in speech and/or music, the brain may automatically apply this expected timing during perceptual processing—referred to by Boltz (1998) as the *imputed velocity hypothesis*. This hypothesis shares many commonalities with more recent Bayesian models of perception, in which prior expectations bias perception towards states that are more likely *a priori* (e.g., Goldreich & Tong, 2013).

The mechanics of sound production may also contain correlations between higher pitches and faster movement. Broze and Huron (2013) noted that lower-pitched instruments tend to be larger, both giving them slower attack transients and requiring greater motor demands from musicians. These factors constrain how quickly many low-pitched instruments can be played. As a result, music written for lower-pitched instruments may tend to be slower than music written for higher instruments. Similarly, Boltz (2017) has suggested that both pitch and movement rate may negatively correlate with body size, as larger people and animals may tend to have lower voices and move more slowly.

Researchers focused on interval timing rather than rhythmic timing also tend to favor explanations based on correlations between size and pitch, but focus instead on perceived stimulus size, rather than speed. Pfeuty and Peretz (2010) noted that if the real-world correlation between smaller size and higher pitch is applied to the size of interonset intervals, then higher rhythms may also be perceived as faster. In contrast, Lake et al. (2014) focused on the perceived size of the tones themselves rather than the intervals between them. They suggested that, due to the Doppler effect, pitch increases as objects approach an observer and decreases as objects recede. As looming objects may be perceived as larger, ascending pitch may cause tones to sound longer—thereby shortening the perceived silent interval between them. Alternatively, looming sounds may be perceived as faster due to greater perceived threat (Neuhoff, 2016).

Although there is good reason to expect correlations between pitch and timing in nature *a priori* due to physical laws, relatively little work has directly tested these correlations in speech and music corpora. One corpus analysis of Western art music has found that higher parts in polyphonic music tend to have more notes, and solos on higher-pitched instruments tend to be faster than those on lower-pitched instruments (Broze & Huron, 2013). However, a large-scale, cross-cultural analysis will be necessary to demonstrate a universal relation between higher pitch and faster timing in human communication. Two recent cross-cultural datasets of speech and song may provide a promising starting point (Mehr et al., 2019; Ozaki et al., 2024), but this is beyond the scope of the present work.

IV	**DV**	**Relation**	**Relevant Studies**	**Chapters**
Pitch	Tempo	$\nearrow$ $\nearrow$ $\nearrow$ $\nearrow$	Boltz (2011, 2017) Collier & Hubbard (1998) Feldstein & Bond (1981) Pfeuty & Peretz (2010)	Ch. 2 Ch. 3
Pitch	ΔTempo	$\nearrow$	Collier & Hubbard (1998)	Ch. 4.3
ΔPitch	Tempo	$\nearrow$ $\nearrow$	Boltz (2011) Collier & Hubbard (1998)	
ΔPitch	ΔTempo	$\nearrow$ $\nearrow$ $\nearrow$ $\nearrow$ $\nearrow$ $\nearrow$	Boasson & Granot (2012) Collier & Hubbard (1998) Gordon & Ataucusi (2021) Herrmann & Johnsrude (2018) Herrmann et al. (2013, 2014) Lake et al. (2014)	
Tempo	Pitch	$\nearrow$ $\nearrow$ $\nearrow$ $\nearrow$	Boltz (2017) Duke et al. (1988) Geringer & Madsen (1984) Madsen et al. (1984)	
Tempo	ΔPitch	?		
ΔTempo	Pitch	?		
ΔTempo	ΔPitch	?		Ch. 4.4 Ch. 5

Table 1.1: Summary of evidence for interactions between pitch and tempo in auditory perception. Chapters that will address each relation are listed in the rightmost column. (Δ = change in feature, $\nearrow$= higher–faster relation, ? = untested)

1.5 Goals of the Present Work

Given the wealth of evidence that pitch influences perceived tempo, but the absence of pitch as a feature among leading models of rhythm perception, I asked what additional data might be useful for informing the incorporation of pitch into these models. Although modeling work is beyond the scope of this thesis, my hope is that the behavioral data presented herein will bring the field closer to an integrated account of pitch and rhythm perception.

The first limitation among existing literature, from a modeling standpoint, is that researchers have typically tested the effects of pitch within a limited range of octaves. The largest pitch differences tested thus far have spanned a two-octave range (Boltz, 2011; Collier & Hubbard, 1998), and the highest pitch tested among those was C6 (1046.4 Hz). Although this level covers the approximate limits of human vocal production, it does not cover the limits of instrumental music—for example, the notes on a piano extend two octaves beyond C6, and the limits of human perception extend at least two octaves further than that. In order to build a more general model of pitch–tempo interactions, we need to understand how pitch influences perceived tempo across the full range of musical experience. Chapter 2 addresses this issue in further detail, and provides an open-access dataset for the benefit of future modeling work. Specifically, I discuss a series of five experiments in which I collected subjective tempo ratings for repeating tones ranging from A2 (110 Hz) to A7 (3520 Hz), in an attempt to map out the shape of illusory tempo effects across a full five-octave range.

It also remains largely unknown whether pitch affects the timing of sensorimotor synchronization in the same manner that it affects subjective ratings of tempo. With the exception of the synchronization tapping experiment by Boasson and Granot

(2012), all other behavioral data concerning the effects of pitch on perceived tempo have been subjective ratings. However, knowing whether the effects of pitch on sensorimotor timing match the effects on subjective tempo can help distinguish whether the influence of pitch occurs upstream of the motor system (e.g., on-line during perceptual processing) or whether it arises through a separate stream, such as being used post hoc as cue during decision making. Therefore, in Chapter 3, I present two sensorimotor experiments designed as a companion piece to the subjective rating studies from Chapter 2. I asked participants to synchronize with the same repeating tones tested in Chapter 2, and then evaluated whether pitch affected their motor timing in the same manner that it affected subjective tempo ratings. The sensorimotor data from these experiments have also been made openly accessible, in the hopes they may prove useful for future modeling work.

The final major gap in the literature that I addressed is whether pitch–timing interactions are bidirectional. As Table 1.1 illustrates, there exists much more evidence of pitch-induced illusory tempo changes than timing-induced illusory pitch changes. Any eventual model of rhythm perception that incorporates pitch will take on substantially different forms depending on whether pitch and timing interact bidirectionally or unidirectionally. If pitch influences rhythm perception without timing influencing pitch perception, then the most appropriate model might be one in which pitch is simply added as a factor to existing predictive coding or neural resonance models. In contrast, if tempo also affects perceived pitch, then the most appropriate model may be an integrated system that jointly estimates both pitch and time. Chapters 4 and 5 address this issue of bidirectionality in detail. Chapter 4 discusses a pair of similar, but inverted, experiments that help establish the bidirectionality of pitch–time interactions. The first experiment tested whether perceived tempo change depends on the pitch of a mistimed tone and/or the pitch of the preceding context; the second tested whether a tone arriving early or late influences its perceived pitch. Chapter 5 then provides an extended analysis of the latter experiment, as well as a follow-up experiment that tests whether timing biases perceived pitch more strongly under conditions of low discriminability.

I then conclude with a General Discussion, summarizing the results of all four data chapters, detailing the implications of this work for future theoretical and computational development, and outlining potential practical applications for interactions between pitch and time in auditory perception.

Chapter 2

Pitch-induced illusory percepts of time

2.1 Preface

In this first collection of experiments, we tested the generalizability of the hypothesis that higher pitches are perceived as faster. Whereas previous experiments have typically compared perceived tempo across a limited range of octaves, our goal was to map out illusory tempo across the broader span of human hearing. We hoped that doing so would facilitate the development of new models of rhythm perception that incorporate the influence of pitch. Unexpectedly, our initial experiment showed that the perception of higher pitches as faster did not generalize across all octaves; rather, perceived tempo increased with pitch to a peak between 440 and 1760 Hz, before decreasing with pitch above that threshold. Our observation of this inverted U-shaped effect prompted us to run a total of four follow-up experiments in an attempt to replicate and explain these results. These five experiments comprise Chapter 2.

We originally designed this study in March 2020 as a two-session experiment. Session 1 would have tested sensorimotor synchronization to tones from different octaves, and Session 2 would have collected the same participant's subjective ratings of the tempos of those same stimuli. The COVID-19 pandemic prompted us to redevelop our subjective rating task for a web-based format and postpone the sensorimotor component until in-person testing resumed. This web-based task became the basis for the present chapter. In April 2022, we were able to implement the postponed sensorimotor task as a separate study, which is reported in Chapter 3.

We submitted the present chapter to *Attention, Perception, & Psychophysics* as a manuscript for publication in January 2023, and submitted our first round of revisions in January 2024. We are currently awaiting the second round of reviews. A preprint of the manuscript associated with this chapter is publicly available on *PsyArXiv* at `https://osf.io/preprints/psyarxiv/6fx87`.

2.2 Introduction

Accurate tempo tracking in auditory perception helps us to direct our attention to critical moments in speech and music (Jones, 1976; Jones & Boltz, 1989; Large & Jones, 1999), and can help us to interpret and communicate emotion (Scherer & Oshinsky, 1977; Scherer, 1986). However, past research suggests that the nontemporal features of an auditory stimulus can distort perceived tempo (e.g., Boltz, 1998, 2011), through a phenomenon we will refer to as *illusory tempo*. In the present study, we investigated how the pitch of an auditory signal biases perceived tempo. In particular, we considered previous observations that people perceive higher-pitched signals to be faster than lower-pitched signals (Boltz, 2011; Collier & Hubbard, 1998), and we tested whether these effects generalize across a wider range of octaves and base tempos than previously evaluated.

2.2.1 Illusory Tempo Effects

Using a paradigm in which participants listened to pairs of melodies and rated their relative tempos, Boltz (1998) showed that melodies with more contour changes or larger pitch intervals were perceived as slower than those with fewer contour changes or smaller pitch intervals. Boltz (2011) later demonstrated additional illusory tempo effects in which melodies played in a higher octave, brighter timbre, or which ascended in pitch or increased in loudness were perceived as faster than those played in a lower octave, duller timbre, or which descended in pitch or loudness. The perception of high-pitched or ascending sequences as speeding up, relative to low or descending sequences, has also been demonstrated in musical scales (Collier & Hubbard, 1998), pitch glides (Gordon & Ataucusi, 2021), and amplitude-modulated tones (Herrmann & Johnsrude, 2018). Consistent with this "higher-equals-faster" illusion, research on perceived interval duration has additionally shown that intervals flanked by at least one high-pitched tone tend to be underestimated, whereas those flanked by at least one low-pitched tone are overestimated (Lake et al., 2014; Pfeuty & Peretz, 2010). In contrast, higher-pitched sounds tend to be perceived as longer than lower-pitched sounds of the same duration (Brigner, 1988; Cohen et al., 1954; Gussenhoven & Zhou, 2013). Meanwhile, research on tempo preferences suggests that listeners prefer higher-pitched melodies to be played faster than lower-pitched melodies, with musical training strengthening this effect (Tamir-Ostrover & Eitan, 2015).

Implicit associations between higher pitch and faster tempo have been established in the literature, as well. A Stroop task by Walker and Smith (1984) found faster reaction times when high tones were paired with the words “fast” and “active” and low tones were paired with “slow” and “passive” than for the reverse pairings. Meanwhile, Eitan and colleagues have shown an association between higher pitch and faster imagined motion in adults (Eitan & Granot, 2006), although not reliably in children (Eitan & Tubul, 2010; Kohn & Eitan, 2009, 2016).

Beyond the domain of music, some evidence suggests that illusory tempo effects also apply to speech. Feldstein and Bond (1981) found that louder or higher-pitched speech was perceived as faster than quieter or lower-pitched speech, with pitch having the strongest effect when speech was quiet. Furthermore, Boltz (2017) found that participants were better able to identify changes in the pitch or tempo of a previously-heard speech sample when both pitch and tempo were changed in the same direction, as compared to when the complementary feature was changed in the opposite direction or left unchanged, suggesting that the brain encodes pitch and tempo in an integrated manner.

Several sensorimotor studies suggest that illusory tempo also affects motor action. Ammirante, Thompson, and Russo (2011) found in a synchronization-continuation tapping task that people tapped slower following tones that changed the pitch contour than following contour-preserving tones, consistent with the perceptual findings from Boltz (1998); however, they also observed faster tapping rates with larger pitch intervals, inverting the previously noted bias to rate melodies with larger intervals as slower. Synchronization tapping studies by Boasson and Granot (2012, 2019) have similarly shown decreases in tapping rates following contour changes, and have also demonstrated that both tapping rate and negative mean asynchrony increase when tapping to sequences that rise in pitch or intensity, consistent with the perceptual ratings from Boltz (2011).

Finally, evidence from Herrmann et al. (2013) suggests that illusory tempo effects may directly influence the dynamics of neural entrainment to a periodic signal. They demonstrated using MEG that the phase lag of stimulus-entrained oscillations in the right auditory cortex shift based on the pitch direction of a frequency-modulated tone. Furthermore, they observed increased intertrial phase coherence when an accelerating or decelerating modulation rate was paired with a congruent (increasing or decreasing) change in pitch, compared to an incongruent one. Together, these results suggest that changes in pitch can alter the relative phase of entrained oscillations.

2.2.2 Origins of Illusory Tempo

Several potential explanations for illusory tempo effects have been proposed. Boltz (1998) has suggested an imputed velocity hypothesis, in which listeners apply expected timing onto stimuli based on learned associations between timing and acoustic features. For instance, melodic phrase boundaries typically coincide with temporal accents; therefore, the imputed velocity hypothesis suggests that listeners will perceive elongated timing at phrase boundaries regardless of whether a temporal accent was actually present (see also Henry & McAuley, 2009, for evidence of imputed velocity in the perceived timing of pitch changes). Likewise, correlations between pitch and tempo have been found in Western music, with higher-pitched voice lines tending to be faster than lower-pitched ones (Broze & Huron, 2013). It might therefore be the case that listeners impute faster timing onto melodies played in higher octaves because higher-pitched melodies are expected to be fast. Furthermore, studies of music visualization have found that adults (Eitan & Granot, 2006) associate rising pitch with spatial acceleration, consistent with rising lines being perceived as faster than falling lines (Boltz, 2011).

Alternatively, Ammirante et al. (2011) has proposed an ideomotor hypothesis, in which perceptual and motor representations of action interact. According to their hypothesis, the perceived tempo of an auditory object can be influenced by movements associated with that sound. For example, perceived slowing at contour changes may originate from the knowledge that bodies must slow down when changing direction. Likewise, faster tapping in response to larger pitch intervals may reflect the idea that one must move faster to travel longer distances in a fixed amount of time (Ammirante et al., 2011).

To explain the effects they observed on neural dynamics, Herrmann et al. (2013) suggested that the brain may recruit oscillators with different natural frequencies to track stimuli with different pitch directions. This hypothesis resembles an earlier proposal by Large (2000) that banks of rhythm-tracking oscillators could integrate pitch information by allowing higher-pitched sounds to preferentially stimulate naturally faster oscillators. In dynamical systems models of rhythm perception, neural oscillators entrain to the frequency at which they are stimulated while remaining attracted to their intrinsic natural frequency (Large & Snyder, 2009; Large et al., 2015). Therefore, if oscillators with different natural frequencies track stimuli with different pitches or pitch directions, pitch may bias neural entrainment towards earlier or later phases, as well as faster or slower tempos (see Kim & Large, 2015 for an analysis of how oscillatory dynamics depend on the relation between the natural frequency of an oscillator and the tempo of its stimulus).

2.2.3 The Present Study

Although there are many nontemporal factors known to influence perceived tempo, the present study focuses specifically on the effect of pitch height. In particular, we seek to address some of the limitations of the previous illusory tempo study conducted by Boltz (2011), in which people judged the relative tempo of differently-pitched melodies. Although they concluded that higher-pitched melodies are perceived as faster than lower-pitched melodies, they compared only one lower register (melodies starting on C3 or 130.8 Hz) to one higher register (melodies starting on C5 or 523.3 Hz). This two-condition design leaves open the question of whether perceived tempo truly increases monotonically with pitch height, or whether it may vary non-monotonically across octaves.

In addition to our concerns over the generalizability of higher pitches being perceived as faster, it must be noted that confounding factors other than pitch height vary between octaves. Figure 2.1 illustrates pitch height alongside two such factors, cochlear sensitivity (related to the perception of loudness; Fletcher & Munson, 1933; International Organization for Standardization [ISO], 2003) and pitch salience (Huron & Parncutt, 1993; Terhardt et al., 1982). Like pitch height, both cochlear sensitivity and pitch salience are greater at C5 than at C3 (Figure 2.1A). However, cochlear sensitivity exhibits substantial nonlinearities across octaves, and pitch salience follows an inverted U-shape that peaks near 700 Hz. Although we agree that pitch height is the most parsimonious explanation of earlier findings, we cannot rule out other potential factors based on existing data.

The present study seeks to address both limitations by testing perceived tempo at pitch heights spanning a five-octave range (Figure 2B), from A2 (110 Hz) to A7

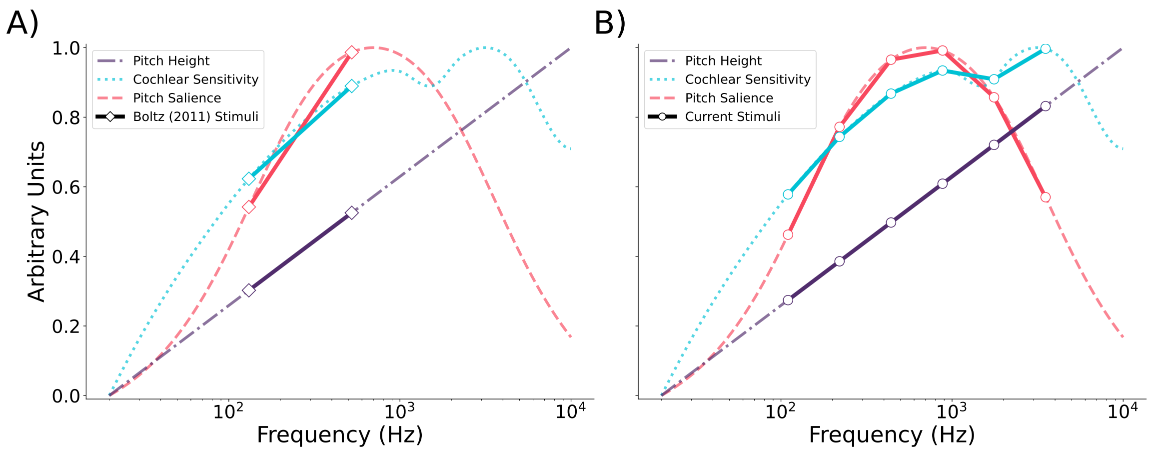

Figure 2.1: An illustration of the benefit of testing pitches spread across several octaves. Background curves illustrate normalized values of pitch height, cochlear sensitivity (based on the equal-loudness contour of ISO 226:2003; International Organization for Standardization, 2003), and pitch salience (based on Terhardt et al., 1982). Points denote A) the fundamental frequencies of the initial tones of melodies used by Boltz (2011), and B) those of the tones used in Experiments 1–4 of the present study. Pitch height, cochlear sensitivity, and pitch salience are all greater in the high-pitched melody condition of Boltz (2011) than in their low-pitched condition (A), and each could account for the increased perceived tempo of their high-pitched melodies. These factors can be disambiguated by sampling across several octaves (B).

(3520 Hz). This procedure allows us to map out the shape of pitch-induced illusory tempo across the majority of the range within which humans perceive pitch most clearly. Mapping the effect across several octaves will confirm whether perceived tempo varies monotonically with pitch height, and will clarify whether pitch-induced illusory tempo is best explained by pitch height or a different factor that varies across octaves. We hypothesized that the relation between pitch and perceived tempo would indeed be monotonic, with higher pitches perceived as faster across a wide range of fundamental frequencies, thereby supporting the generalizability of the higher-equals-faster illusion.

2.3 Experiments 1–3

Our first three experiments sought to determine the shape of the pitch-induced illusory tempo effect across a wide pitch spectrum and across a broad range of stimulus tempos. We did so using a relative tempo judgment paradigm in which participants heard a standard metronome sequence followed by an isochronous repeating tone, and then rated on a sliding scale how fast the repeating tone was relative to the standard. The repeating tone varied both in pitch and tempo across trials. We further investigated whether pitch-induced illusory tempo can be resisted through sensorimotor synchrony, as prior literature suggests movement can improve temporal judgments in some cases (Butler & Trainor, 2015; Manning & Schutz, 2013, 2016) but not others (London, Thompson, Burger, Hildreth, & Toiviainen, 2019). Specifically, some participants were instructed to tap in time with the stimuli while others were instructed to minimize their movements while listening.

Procedurally, these three experiments were nearly identical, but each used a different stimulus tone designs. Experiment 1 presented participants with 200 ms piano tones, but these tones varied naturalistically in their sustain, with higher tones tending to have sharper decays than lower tones. Experiment 2 reduced variability in the tones' amplitude envelopes between octaves by presenting truncated, 50 ms piano tones. Experiment 3 returned to the longer, 200 ms tone design of Experiment 1 but guaranteed that tones sustained for the full 200 ms regardless of pitch. For conciseness, due to the similarity between these experiments, and in order to increase our power to detect interaction effects between pitch, stimulus tempo, and tapping behavior, we have pooled the data from these three experiments into a single analysis.

2.3.1 Methods

Participants

Across Experiments 1–3, a total of 193 undergraduate students (58 male, 135 female) from McMaster University participated for course credit. The participant counts for each of the three experiments were 44 (25 female), 78 (51 female), and 71 (59 female), respectively. Ages ranged from 16–31 years, with mean ages of 18.9 ($SD = 1.3$), 19.8 ($SD = 2.1$), and 18.2 ($SD = 1.3$) years across the three experiments, respectively. Within each experiment, participants were randomly assigned to either the tapping or non-tapping condition, with a total of 97 (71 female) assigned to the tapping condition and 96 (64 female) assigned to the non-tapping condition. We conducted the experiments sequentially between April and September 2020, and individuals were not permitted to participate in more than one of the three experiments. All participants completed the experiments online due to the COVID-19 pandemic.

An additional 16 participants completed the study but were excluded using the following criteria. Ten participants were excluded for treating the slider as a discrete response scale. We classified participants as discrete-responders if they responded with a rating of 0, 50, or 100 on at least 75 out of 90 trials. We excluded an additional six participants for having a low or negative correlation ($r < .5$) between their responses and the true relative tempo of the stimulus (see Response Scoring for the calculation of ground-truth tempo ratings).

Data Availability

We have made all data, code, and stimuli from all five experiments publicly available on the Open Science Framework at `https://osf.io/85cyx/`, as well as on GitHub at `https://github.com/jpazdera/IllusoryTempo`.

Materials

In Experiments 1 and 2, we generated the seven unique tones used in the study (A2, A3, A4, A5, A6, A7, and D♯5), as well as the sound of the metronome, using the music composition software MuseScore. The tick of the metronome was a simulated wood block sound which lasted 50 ms in duration and the tones were piano tones which varied in duration prior to editing. We used Python (version 3.7) to convert all eight sounds from stereo to mono, truncate all piano tones to either a 200 ms (Experiment 1) or 50 ms duration (Experiment 2), and apply a 10 ms linear fade-out to each tone. In Experiment 3, we used the same metronome as Experiments 1 and 2, but generated a new set of piano tones using Sonic Pi (version 3.2) to ensure that all tones sustained longer than 200 ms. We then used Python to convert the tones from stereo to mono, truncate them to a 200 ms duration, and apply a 25 ms linear fade-out.

We additionally equalized the perceived loudness of the tones in each experiment to limit the potential effects of loudness differences on perceived tempo (Boltz, 2011). In Experiment 1, we equalized the baseline loudness of all tones to -16 LKFS using the pyloudnorm package for Python (Steinmetz, 2019), which implements ITU-R BS.1770-4 (International Telecommunication Union [ITU], 2017). In Experiments 2 and 3, we instead used the `acousticLoudness` function in MATLAB (version R2020a) to equalize the perceived loudness of all tones according to ISO 532-1:2017 (ISO, 2017). For each experiment, we then used the audio editing software Audacity to generate louder (+3 dB) and softer (−3 dB) variants of each tone for a total of 21 piano tone stimuli.

To ensure precise inter-onset timing, we pre-generated all sound sequences using Python, and played them back as WAV files during the experiment.

Apparatus

We used the JavaScript library jsPsych (de Leeuw, 2015) to implement our stimulus presentation and response collection, and conducted the experiment online through Pavlovia (https://pavlovia.org). We performed all analyses using Python (version 3.10) and R (version 4.3).

Design

The study followed a mixed 2 (Tap Instruction) × 6 (Pitch Height) × 15 (Stimulus Tempo) design. Participants assigned to the tapping condition were instructed to tap the J or F key with their dominant hand in synchrony with both the metronome and repeating tone on each trial. Participants in the non-tapping condition were instead instructed to listen to both sequences while keeping their movements to a minimum. In Experiments 2 and 3, reminder text was also displayed ("Remember to [tap along/avoid moving] while listening.") while the metronome and repeating tone played on each trial. Each trial used one of six pitch heights for its repeating tone: A2 (110 Hz), A3 (220 Hz), A4 (440 Hz), A5 (880 Hz), A6 (1760 Hz), or A7 (3520 Hz). The interonset interval (IOI) of the repeating tone was set to one of 15 log-spaced values on each trial: 1000, 918, 843, 774, 710, 652, 599, 550, 504, 463, 425, 390, 358, 329, or 302 ms. These intervals correspond to the following tempos in beats per minute (BPM): 60.0, 65.4, 71.2, 77.5, 84.5, 92.0, 100.2, 109.1, 119.0, 129.6, 141.2, 153.8, 167.6, 182.4, and 199.7. We chose to center the timing conditions around 550 ms (109.1 BPM), as this tempo falls approximately where rhythm perception exhibits the greatest sensitivity and least bias (e.g., Drake & Botte, 1993; Friberg & Sundberg, 1995; Vos, van Assen, & Fraňek, 1997). We then selected an overall range of between 60 and 200 BPM to provide good coverage of the tempos most commonly observed in music, while keeping the rate solidly within the limits of human performance (Repp, 2006). For certain analyses, as well as during trial randomization, these 15 stimulus tempos were binned into five "tempo ranges" based on their quintile, such that 1000, 918, and 843 ms comprised the slowest tempo range, 774, 710, and 652 ms comprised the second slowest range, and so forth. On practice trials the pitch of the repeating tone was D♯5 (622.25 Hz, the midpoint of A2 and A7) and the interonset interval on each trial was one of 741, 550, and 407 ms.

Given the variability in online participants' audio equipment—which may have affected the relative loudness of different pitches depending on their specific hardware—

we additionally manipulated tone loudness as a control variable, in order to test whether small differences in intensity may affect perceived tempo. The repeating tone on each trial was either played at its baseline loudness, 3 dB louder than baseline, or 3 dB quieter than baseline.

Procedure

Before receiving instructions for the main task, participants completed a headphone test based on that of Woods, Siegel, Traer, and McDermott (2017). The test consisted of six trials in which participants listened to three 200 Hz pure tones and rated which was quietest. Two of the tones were played with equal loudness, while the target was played at 6 dB below that level. However, one of the equal-loudness tones was phase-reversed between stereo channels, making it sound quieter than the other when played over speakers. Consequently, participants wearing headphones reliably identify the target tone as quietest, while those listening through speakers do not (Woods et al., 2017).

Following the headphone test, we measured each participant's spontaneous motor tempo (McAuley, Jones, Holub, Johnston, & Miller, 2006) by asking them to rest their palms on the surface in front of them and tap with the index finger of their dominant hand on the J or F key at the rate that felt most natural and comfortable to them. After 20 (Experiment 1) or 15 (Experiments 2 and 3) seconds of tapping, participants were stopped and given the main task instructions.

Each trial featured a relative tempo judgment task in which participants listened to five ticks of a metronome followed by five repetitions of a piano tone. The metronome was identical on all trials and always played with an interonset interval of

550 ms, whereas the tone varied across trials in pitch, interonset interval, and loudness. An interval of 1800 ms of silence separated the metronome from the repeating tone on all trials. Following the repeating tone, participants were given a continuous slider and asked "How fast was the repeating tone relative to the metronome?" The slider's scale ranged from "Half as Fast" (a score of 0) on the left to "Twice as Fast" (a score of 100) on the right, with "Equal Rates" (a score of 50) labeled in the center. Participants could only see the descriptive labels, and were not able to view the numeric values assigned to their responses. Participants were not limited in how long they could take to respond, and the next trial began 1500 ms after they submitted a response.

Each participant completed three practice trials, as well as one trial under each of the 90 unique pitch × tempo pairings. Among the three presentations of each pitch within each of the five "tempo ranges," one trial was played at each of the three loudness levels. Trials were arranged into three blocks of 30 and were randomly ordered, with the constraint that each pitch was paired with one tempo from each of the five tempo ranges exactly once per block. Additionally, consecutive trials within a block always differed in both pitch and tempo range. Participants were allowed self-paced breaks between blocks.

Response Scoring

Deriving Perceived Tempo from Raw Ratings To facilitate interpretation of the 0–100 ratings on the response scale, we first defined ground-truth ratings for all stimulus tempos. These ground-truth ratings were based on the assumption that—given the labeling of the response scale—a rating of 50 should correspond to a stimulus

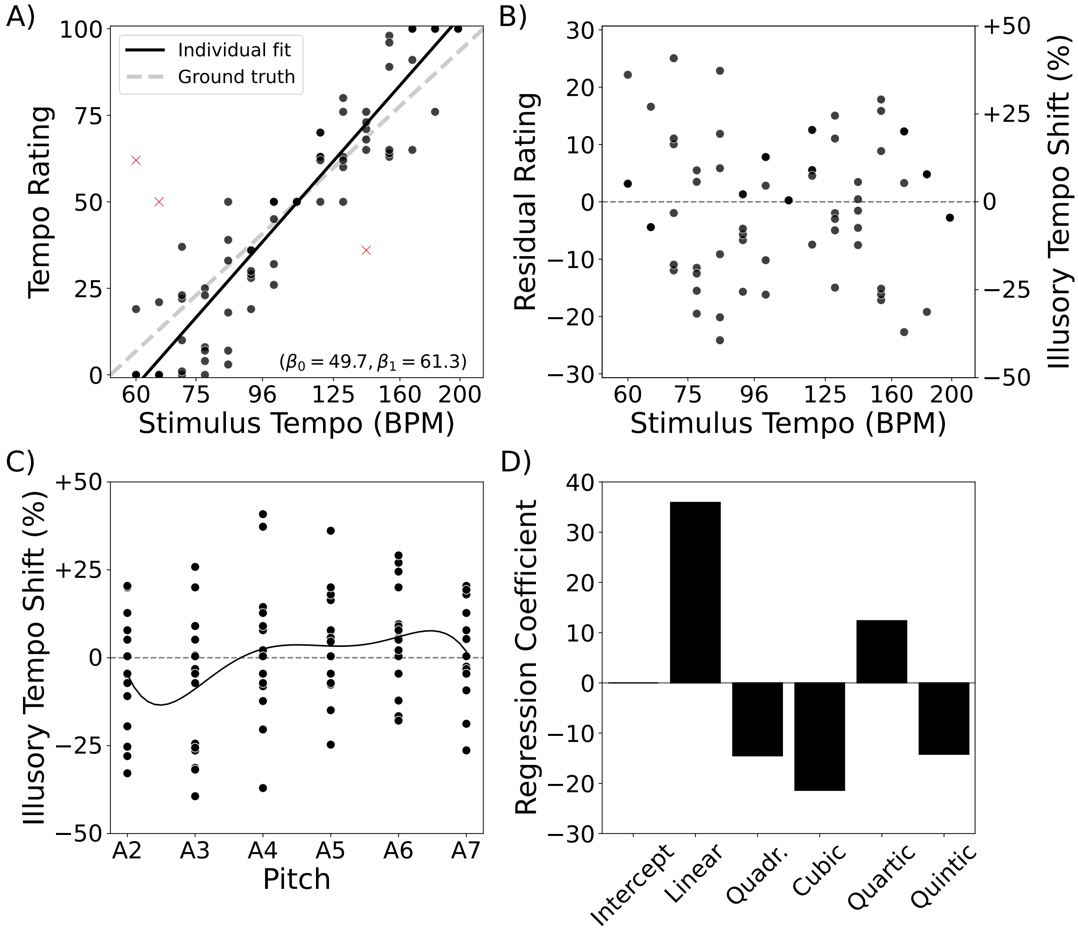

Figure 2.2: Data processing steps for one example participant. **A)** A line is fit to the participant's raw tempo ratings (black dots) as a function of the log of the stimulus tempo (Equation 2.3.2). After an initial fit, outlier ratings (red x's) are identified and excluded, and the line is refit. Ground-truth relative tempo, based on the labeling of the response scale, is shown for comparison (dashed line; Equation 2.3.1). **B)** The effect of stimulus tempo is regressed out by subtracting the expected rating for each trial from the participant's response, giving a set of residual tempo ratings (Equation 2.3.4). Residual tempo ratings are then converted to a percent illusory tempo shift by dividing the residuals by the slope of the regression line. **C)** A polynomial regression is fit to predict the participant's illusory tempo shift as a function of pitch height. **D)** The subject-level regression coefficients are extracted, and their distributions across participants are later analyzed to identify the shape of the illusory tempo curve. Shown here are the regression coefficients from one participant. Note that the polynomial's intercept equals zero due to having already regressed out the average rating using Equation 2.3.4.

tempo equal to that of the metronome (i.e., 109.1 BPM or a 550 ms IOI), and a rating increase of 50 points should correspond to a doubling of the tempo (i.e., a halving of the IOI). This mapping assumes that tempo is perceived on a log-linear scale, which can be described in our study by the linear equation:

$$r_{true} = 50 + 50 \log_2 \left(\frac{t}{t_{ref}} \right) \tag{2.3.1}$$

where r_{true} is the ground-truth rating for a given relative tempo, t/t_{ref}, where t is the stimulus tempo in BPM and t_{ref} is the tempo of the metronome in BPM.[1]

However, we did not expect all participants to map relative tempo onto the response scale in a manner identical to the ground truth. For example, some individuals may use the full sliding scale liberally while others may give conservative judgments that deviate less from the midpoint of the scale. We therefore sought to account for individual differences in response scale usage when scoring our response data, before considering the effect of pitch on these ratings. To do so, we fit separate linear models to each participant's responses, predicting their tempo ratings as a function of the log of the relative tempo of the stimulus:

$$\hat{r}_i = \beta_0 + \beta_1 \log_2 \left(\frac{t_i}{t_{ref}} \right) \tag{2.3.2}$$

where t_{ref} is the tempo of the metronome, t_i is the stimulus tempo on a given trial i, and $\hat{r}_i$ is the predicted rating the participant would assign to that stimulus tempo. The intercept, β_0, is their expected rating when the stimulus has the same tempo as

[1]To understand this equation, note that when the stimulus tempo is equal to that of the metronome, the relative tempo equals 1, $\log_2(1) = 0$, and thus $r_{true} = 50$, the intercept of the line. Doubling the stimulus tempo, t, doubles the relative tempo, thereby increasing $\log_2(t/t_{ref})$ by 1 and increasing r_{true} by 50, the slope of the line.

the metronome, and the slope, β_1, describes how many points their rating is expected to change each time the stimulus tempo doubles.

After an initial fit of each participant's linear model, we marked trials with Cook's distance greater than 4/90 (where 90 is the trial count) as outliers, then refit the model with outliers excluded. The outlier-excluded slopes and intercepts were used for the remainder of our analyses, and the outlier trials (932 total, 5.4% of trials) were excluded from all further analysis. Figure 2.2A shows the tempo ratings from one example participant, and illustrates the linear model mapping stimulus tempos to their associated ratings on the response scale. Also shown is the ground truth line from Equation 2.3.1, for comparison.

Note that Equation 2.3.2 can be inverted to estimate the perceived tempo, $\hat{t}_i$, associated with any given rating, r_i:

$$\hat{t}_i = t_{ref} * 2^{(r_i - \beta_0)/\beta_1} \tag{2.3.3}$$

where t_{ref} is the tempo of the metronome in BPM.

Quantifying Illusory Tempo Shift To address the question of how pitch influences perceived tempo, we analyzed whether participants' responses to tones of each pitch differed systematically from the response that would be expected based solely on the tempo of those tones. We did so by regressing out the expected tempo ratings (Equation 2.3.2) from the participants' response values. We define the residual tempo rating, ρ_i, for trial i as the difference between the participant's response on that trial,

r_i, and the predicted rating, $\hat{r}_i$, for stimuli of that trial's tempo:

$$\rho_i = r_i - \hat{r}_i \tag{2.3.4}$$

Thus, a positive residual tempo rating indicates that the participant rated the repeating tone on that trial as faster than they would typically rate a stimulus of that tempo, whereas a negative residual tempo rating indicates that the participant rated the repeating tone as slower than would be expected. Figure 2.2B illustrates residual tempo ratings for one participant's trials after regressing out the stimulus tempo. Note that the average residual rating within a participant will always equal zero, as we have regressed out the participant's average rating.

To translate each participant's residual tempo ratings into percent changes in perceived tempo, we divided the residuals by that person's regression slope, β_1, from Equation 2.3.2. As β_1 describes the number of rating points associated with a doubling in perceived tempo, our illusory tempo shifts should be understood as a percent of a *doubling* in BPM, not a percent change in raw BPM.[2] This scaling follows the overarching assumption in our study that tempo is perceived on a log-linear scale. We thus define the illusory tempo shift, τ_i, on a given trial, i, as:

$$\tau_i = 100\% * \frac{\rho_i}{\beta_1} \tag{2.3.5}$$

The two y-axes of Figure 2.2B illustrate the mapping of residual ratings to illusory tempo shift, based on the example participant's regression slope of $\beta_1 = 61.3$.

[2] While a doubling of perceived tempo would correspond to both a 100% increase in BPM and an illusory tempo shift of +100%, a halving of perceived tempo would correspond to a 50% reduction in BPM, but an illusory tempo shift of −100%.

Note that the above formulation of illusory tempo shift is equivalent to the $\log_2$-percent change in perceived tempo (see Törnqvist, Vartia, & Vartia, 1985 for a discussion of log-percents):

$$\tau_i = 100\% * \log_2\left(\frac{\hat{t}_i}{t_i}\right) \tag{2.3.6}$$

where t_i is the stimulus tempo on trial i, and $\hat{t}_i$ is the perceived tempo associated with the participant's rating on that trial (see Equation 2.3.3).

Data Analysis

Comparing Subjective Ratings to the Ground Truth As a preliminary analysis of the overall performance of participants in our task, we compared the slopes and intercepts of the subject-level linear models of tempo ratings as a function of stimulus tempo (Equation 2.3.2) against the ground-truth model of relative tempo (Equation 2.3.1). We did so using a one-sample Hotelling's T^2 test (a multivariate t-test) with the intercepts and slopes both as outcome variables. This test assessed whether the average subject-level model differed from the ground-truth in terms of intercept ($\beta_0 = 50$) and/or slope ($\beta_1 = 50$). We then conducted post hoc univariate t-testing to determine which of these two model parameters significantly differed from the ground truth, and used the Holm–Bonferroni technique to correct for multiple comparisons.

Assessing Pitch-Induced Illusory Tempo To investigate how perceived tempo varied across octaves, we used polynomial regression as a method of quantifying the pattern of illusory tempo shift across our six levels of pitch height. We used orthogonal polynomials in all of these models to eliminate multicollinearity between regression coefficients. In all cases where we analyzed a polynomial regression model, we first performed multivariate statistics (Hotelling's T^2 or MANOVA) with all model coefficients of interest as outcome variables. We did so as an omnibus test of whether the average of all subject-level models differed as a whole from a null model (in one-sample tests), or whether the average model differed between conditions (in two-sample tests). Then, in any case where a multivariate analysis produced a significant result, we conducted post hoc univariate t-testing to determine which specific model coefficients accounted for the significant effect. We corrected for multiple comparisons in each post hoc analysis using the Holm–Bonferroni technique.

Main Effect of Pitch To evaluate the main effect of pitch height on perceived tempo, we analyzed how illusory tempo shift (Equation 2.3.5) differed across octaves. To quantify the within-subject shape of the illusory tempo effect, we fit separate fifth-order polynomial regression models to each person's data, predicting illusory tempo shift as a function of pitch height (coded as integers from 2 to 7, prior to making the polynomial factors orthogonal). Figure 2.2C illustrates the regression model for one example participant, and Figure 2.2D shows the coefficients of this model. As every participant's average residual tempo rating equaled zero (see Response Scoring), the intercept of the illusory tempo polynomial also always equaled zero. Therefore, we analyzed only the distribution of slopes across participants. Specifically, we conducted a one-sample Hotelling's T^2 test with the linear, quadratic, cubic, quartic, and quintic slopes of the polynomial regression models as the outcome variables. This test indicated whether the average model of pitch-induced illusory tempo shift differed from a null model in which all slopes were zero.

Effect of Pitch Across Stimulus Tempos After quantifying the shape of the illusory tempo curve across octaves, we tested whether the effect of pitch height differed between the five stimulus tempo ranges. To do so, we fit five separate second-order polynomial regression models within each participant—one to their illusory tempo shift data from each tempo range. These models were of an order equal to the highest significant model order determined during our analysis of the main effect of pitch. This procedure gave linear and quadratic slopes describing the shape of the illusory tempo effect within each of the five stimulus tempo ranges. We then conducted a repeated-measures MANOVA with the five levels of tempo range as the independent variable and the linear and quadratic slopes as dependent variables. Multicollinearity was avoided through the use of orthogonal polynomials, as noted above. Multivariate normality was inspected graphically, which suggested higher than normal kurtosis but a symmetrical distribution, to which MANOVA should be robust. A Box's M test indicated significant heterogeneity of covariances; however, our balanced design should allow robustness to such a violation, as well.

Effect of Tapping on Illusory Tempo To determine whether synchronous tapping attenuated the illusory tempo effect, we fit a second-order polynomial regression model to each participant's illusory tempo shifts as a function of pitch height, while excluding trials on which participants who were instructed to tap failed to do so (27.6% of trials). We then conducted an independent-samples Hotelling's T^2 test to determine whether the linear and quadratic slopes differed between participants in the two tapping conditions.

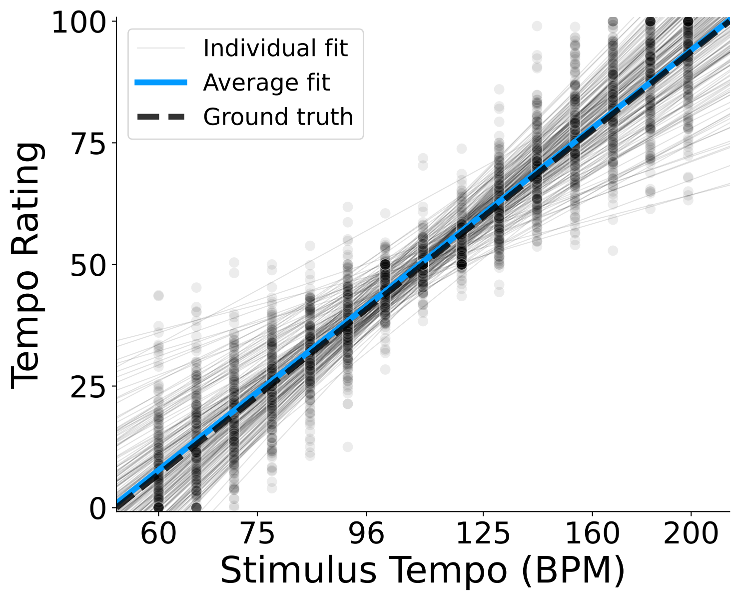

Figure 2.3: Points show participants' average ratings for each stimulus tempo in Experiments 1–3. Ground-truth ratings (Equation 2.3.1) are indicated by the dashed line. Linear regression lines for individual participants (Equation 2.3.2) appear as thin gray lines, and the average of all regression lines appears in blue.

Effect of Loudness on Perceived Tempo Finally, we tested whether our control variable, loudness, affected the perceived tempo of the repeating tones. We did so using a repeated-measures ANOVA to identify any differences in average illusory tempo shift between tones played 3 dB above, below, or at baseline. A significant deviation from sphericity was detected via Mauchly's test, and Huynh-Feldt correction was applied accordingly ($\varepsilon = .957$).

2.3.2 Results

Subjective Ratings versus Ground Truth

Figure 2.3 illustrates subject-level tempo ratings as a function of the repeating tone's tempo, in comparison to the theoretical ground-truth ratings (dashed line; Equation 2.3.1). Also shown is each participant's individual regression line (thin gray lines; Equation 2.3.2), used in deriving their residual tempo ratings and illusory tempo shift scores. The average of all individual fits is shown, as well (blue line). The average fit ($\beta_0 = 50.59$, $\beta_1 = 49.70$) differed significantly from the ground truth relative tempo, $F(2, 191) = 3.50$, $p = .032$, $\eta_p^2 = .035$, and post hoc testing revealed that the average intercept was significantly greater than the ground truth of 50, $t(192) = 2.62$, $p = .009$, $d = 0.189$, $CI = [50.15, 51.04]$. The average slope ($M = 49.70$), however, did not differ significantly from the ground truth, $t(192) = -0.40, p = .693, d = 0.029$, $CI = [48.20, 51.20]$. Though significant, the small difference between participants' average behavior and the true relative tempo suggests that they were generally quite accurate at the task, but tended to rate the repeating tone as slightly faster than it actually was.

Pitch-Induced Illusory Tempo

Our primary research question was how pitch height would affect participants' tempo ratings across a wider frequency range than previously studied. Figure 2.4A illustrates the percent illusory tempo shift observed for repeating tones at each pitch height. Illusory tempo followed an inverted U-shaped pattern, in which the lowest (A2; 110 Hz) and highest (A7; 3520 Hz) tones were rated as slowest, whereas the middlemost tones (A4 and A5; 440 and 880 Hz) were rated as fastest. Figure 2.4B shows the average slope coefficients of the polynomial regression models fit to each participant's illusory tempo data. A multivariate analysis of these coefficients revealed that pitch exerted a large and significant overall effect on perceived tempo, $F(5, 188) = 10.51$, $p < .001$, $\eta_p^2 = .218$. Post hoc testing indicated a marginal positive linear effect,

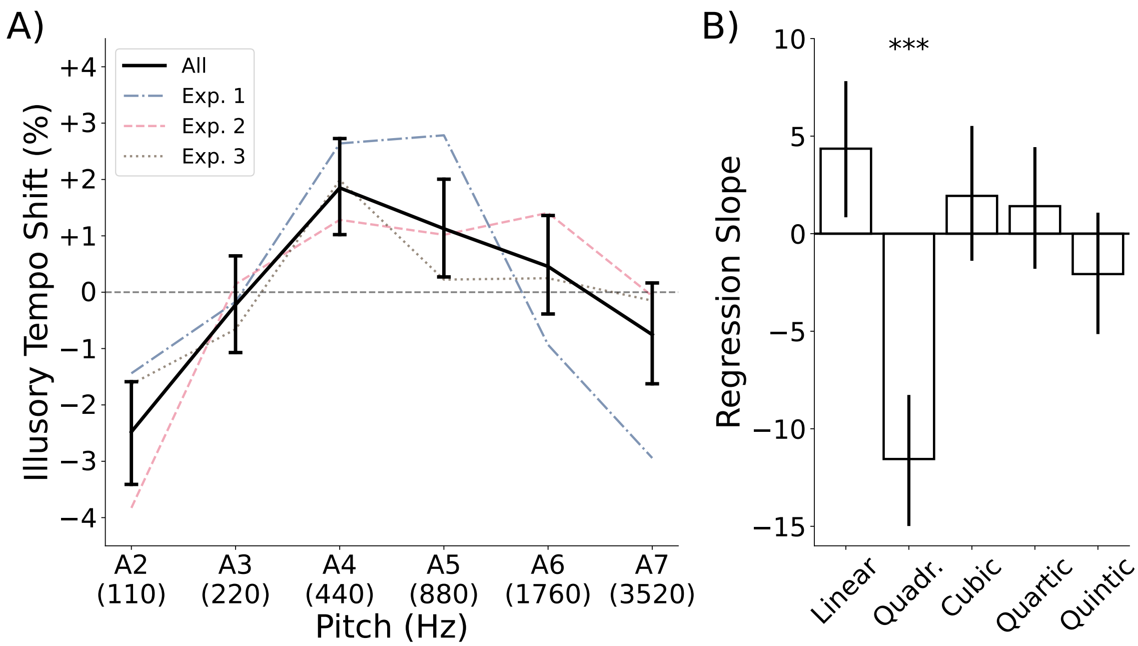

Figure 2.4: **A)** Average illusory tempo shifts across participants for repeating tones of each pitch height in Experiments 1–3. Perceived tempo followed an inverted U-shaped pattern, with extremely high and low tones rated as slower than middle tones. **B)** Average slope coefficients across subject-level polynomial regression models, predicting illusory tempo shift as a function of pitch height. Error bars in both panels indicate 95% confidence intervals.

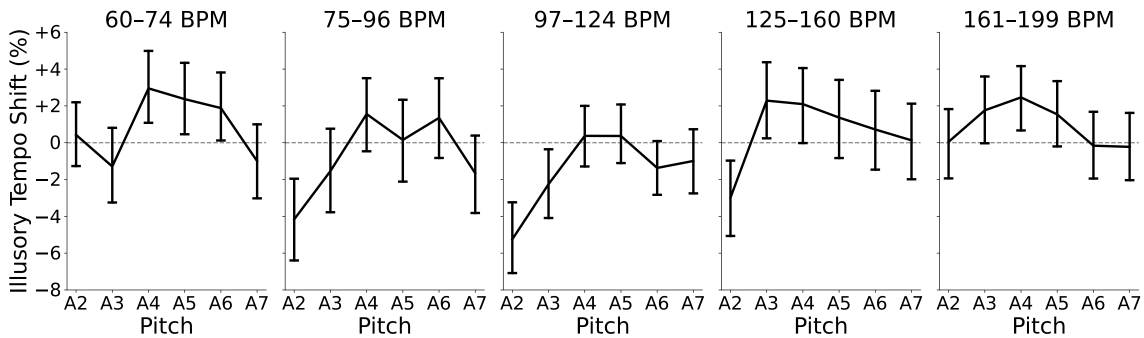

Figure 2.5: Average illusory tempo shifts across participants for repeating tones of each pitch height in Experiments 1–3, conditional on the true stimulus tempo. Error bars indicate within-subject 95% confidence intervals. Regardless of tempo range, the illusory effect of pitch consistently followed an inverted U-shaped pattern.

$t(192) = 2.40$, $p = .017$, $d = 0.173$, and a significant negative quadratic effect of pitch height, $t(192) = -6.87$, $p < .001$, $d = 0.494$. The cubic, $t(192) = 1.10$, $p = .275$, $d = 0.079$, quartic, $t(192) = 0.49$, $p = .622$, $d = 0.036$, and quintic effects of pitch height, $t(192) = -1.45$, $p = .149$, $d = 0.104$, were all nonsignificant. A combined positive linear and negative quadratic effect is consistent with the slightly asymmetrical inverted-U shaped curve seen in Figure 2.4A.

Next, we considered whether the effect of pitch might differ across stimulus tempo ranges. Figure 2.5 shows illusory tempo shift for each pitch, conditional on the tempo range of the repeating tone. The linear and quadratic slopes of the illusory tempo curve did not significantly differ between tempo ranges, Wilk's $\Lambda = 0.983$, $F(8, 1534) = 1.65$, $p = .107$, $\eta_p^2 = .009$, indicating a lack of interaction between pitch and stimulus tempo. Pitch influenced the perceived tempo of the repeating tone similarly, regardless of the tempo at which it was played.

Finally, we considered whether synchronous movement may attenuate the tempo-

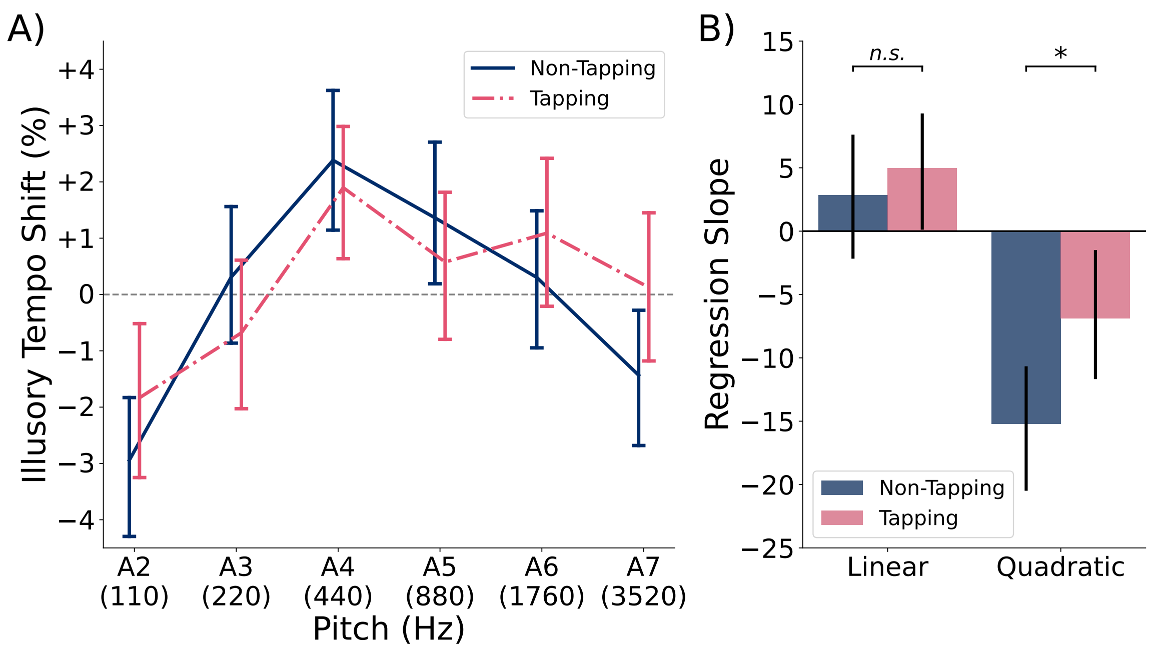

Figure 2.6: **A)** Illusory tempo shifts for repeating tones of each pitch height, conditional on whether the participant was instructed to tap. Averages across participants instructed not to tap are shown in blue, while averages across participants instructed to tap are shown in pink. **B)** Average slope coefficients across subject-level polynomial regression models, predicting illusory tempo shift as a function of pitch height. Error bars in both panels indicate 95% confidence intervals. Synchronous tapping slightly attenuated the negative quadratic effect of pitch on perceived tempo.

biasing effect of pitch. Figure 2.6A illustrates the effect of pitch among participants instructed to tap, and among those instructed to minimize their movements. The polynomial regression slopes for each condition are shown in Figure 2.6B. We observed a small difference between the illusory tempo curve of participants instructed to tap and those instructed not to, $F(2, 181) = 3.28$, $p = .040$, $\eta_p^2 = .035$. The linear effect of pitch was not found to differ significantly between conditions, $t(182) = -0.64, p = .526, d = 0.089$, but the quadratic effect of pitch was slightly weaker among participants in the tapping condition, $t(182) = -2.29, p = .023, d = 0.330$.

Loudness Control

Although we normalized all of our tones to be equally loud, our study was conducted online, where certain pitches may have sounded louder than others depending on participants' hardware and individual hearing loss. To assess whether subtle loudness differences between pitches may have affected participants' tempo ratings, we varied loudness across trials within a 6 dB range (see Methods). Loudness did not significantly affect tempo ratings, $F(1.91, 367.66) = 0.79$, $p = .449$, $\eta_p^2 = .004$. Average illusory tempo shifts for tones played at 3 dB below baseline, at baseline, and 3 dB above baseline were +0.34% ($SD = 3.88$), −0.17% ($SD = 3.29$), and −0.17% ($SD = 4.04$), respectively. These small, nonsignificant differences in perceived tempo across a 6 dB range suggest that variability in the loudness of tones from different octaves is unlikely to have produced the effects of pitch that we observed.

2.3.3 Discussion

Experiments 1–3 investigated the effect of pitch height on perceived tempo across tones ranging from A2 (110 Hz) to A7 (3520 Hz). They revealed an inverted U-shaped effect, such that perceived tempo increased with pitch from A2 to A4 (440 Hz), then decreased with pitch between A4 and A7. This pattern was slightly asymmetrical, with a steeper upward than downward slope, characterized by a significant negative quadratic effect and a marginal positive linear effect. These results help to refine previous suggestions that perceived tempo increases monotonically with pitch (Boltz, 2011; Collier & Hubbard, 1998), and highlight the value of testing effects across several levels of continuous independent factors. Our results are consistent with the findings of Boltz (2011) within the range of fundamental frequencies tested in their study (see Figure 2.1). Specifically, they compared the perceived tempo of melodies starting on C3 or C5 (130.8 Hz or 523.3 Hz; three semitones above A2 and A4, respectively) and found that participants perceived the higher-pitched melodies as faster. This design places their high-pitched melodies close to the point of fastest perceived tempo in our illusory tempo curve and their low-pitched melodies at a point where tempo was perceived to be slower in our illusory tempo curve (see Figure 2.4). However, the results of our Experiments 1–3 suggest a non-monotonic relation between pitch and perceived tempo when the full frequency range we tested is considered.

We also tested whether the effect of pitch varied across stimulus tempos, as well as whether sensorimotor synchronization would attenuate pitch-induced biases. Our results suggest that the effect of pitch on perceived tempo is similar regardless of stimulus tempo. Across five tempo ranges with interonset intervals between 1000 ms (60.0 BPM) and 302 ms (199.7 BPM), we consistently found a similar inverted U-shaped effect (see Figure 2.5). Regarding the effects of sensorimotor synchronization, we observed a small, but significant reduction in the quadratic effect of pitch on perceived tempo among participants who were instructed to tap in synchrony with the stimuli (see Figure 2.6). In conjunction with previous research demonstrating illusory tempo effects on tapping rates during both synchronization (Boasson & Granot, 2012, 2019) and continuation tapping paradigms (Ammirante et al., 2011), we believe it is unlikely that pitch-induced timing biases can be entirely resisted through motor engagement, though weak attenuation may be possible.

As the observed inverted U-shaped effect of pitch was unexpected, we conducted two follow-up experiments in which we evaluated whether specific confounding variables may have driven the negative quadratic effect. One explanation is that we may be observing an effect of pitch salience, instead of (or in addition to) pitch height. Indeed, Figure 2.4 somewhat resembles the pitch salience curve defined by Terhardt et al. (1982). Their model of pitch salience peaks at 700 Hz—approximately 4 semitones below A5 (880 Hz)—and declines symmetrically above and below this point (see Figure 2.1). Furthermore, the effect of pitch height in Experiments 1–3 may have been confounded with an effect of pitch salience because the piano tones we used contained some audible low-frequency content, deriving from the mechanics of the piano. This low-frequency content was audible as a subtle knocking sound that was particularly noticeable in our A6 and A7 tones—possibly due to the pitch of these tones being less salient relative to the knocking (Terhardt et al., 1982). If participants attended on some trials to the low-pitched knocking sound rather than to the highest-pitched piano tones, we might expect them to rate the highest tones as slower than they otherwise would. We therefore conducted Experiment 4 specifically to address whether the inverted U-shaped effect persists when participants listen to controlled, synthetic tones that lack low-frequency content.

An alternative possibility is that a midpoint effect influenced participants' tempo ratings, such that they perceived tones in the middle of the pitch range to be fastest. The fact that tempo ratings peaked between the middlemost pitches in our experiments (A4 and A5) supports this explanation. We conducted Experiment 5 to address the possibility of a midpoint effect by presenting each participant with six tones from only one of the two halves of our pitch range.

2.4 Experiment 4

Experiment 4 addressed the possibility that the unexpected negative quadratic effect of pitch may be specific to the piano tones we tested in Experiments 1–3. To investigate whether the low-frequency content in our piano tones led to the decrease in perceived tempo at our highest octaves, we presented participants with the piano tones from Experiment 3 on some trials and controlled, synthetic complex tones on others. If the quadratic component depends on low-frequency content becoming more salient (Terhardt et al., 1982) relative to the tone at high octaves, the downturn in perceived tempo above 700 Hz should be eliminated for synthetic tones.

2.4.1 Methods

Participants

Seventy-seven undergraduate students (27 male, 50 female) from McMaster University completed the experiment for course credit. Ages ranged from 18–24 years ($M = 18.6$,

$SD = 1.2$). Participants were randomly assigned to hear piano tones on the first block ($N = 36$; 21 female) or synthetic tones on the first block ($N = 41$; 29 female). All participants completed the task online between February and April 2021 due to the COVID-19 pandemic. An additional three participants completed the experiment but were excluded due to low correlation ($r < .5$) between their responses and the true relative tempo (see Response Scoring).

Materials

The metronome was identical to that used in Experiments 1–3 and the piano tones were identical to those used in Experiment 3. The synthetic tones used in Experiment 4 were complex tones created in Python (version 3.8) by summing four sine waves with random phase, including the fundamental frequency and the first three overtones with relative amplitudes defined by a slope of -6 dB/octave. The tones were 200 ms in duration and consisted of a 5 ms linear rise, a 170 ms sustain, and a 25 ms linear decay, approximately matching the amplitude envelopes of the piano tones. We used Audacity's loudness normalization function, which is based on recommendation ITU-R BS.1770-4 (ITU, 2017), to balance all tones to −17, −14, and −11 LUFS for the soft, normal, and loud versions of each tone, respectively.[3] All sound sequences were again pre-generated using Python, and were played back as WAV files during the experiment.

Apparatus

We conducted the experiment online using the same apparatus as Experiments 1–3.

[3]We set loudness to be similar to other web-based content, which most streaming platforms normalize to −14 LUFS. A difference of ±3 LUFS is equivalent to the ±3 dB difference between loudness conditions in Experiments 1–3.

Design

Experiment 4 followed a 2 (Timbre) × 6 (Pitch Height) × 15 (Stimulus Tempo) fully within-subjects design. The repeating tone on each trial was either a synthetic complex tone or one of the piano tones used in Experiment 3. The pitch heights and stimulus tempos of the tones were identical to those used in Experiments 1–3, and tempos were again binned into the same five tempo ranges. Tone loudness was again included as a control variable, with the repeating tone either being played at its baseline loudness, at 3 dB above its baseline, or at 3 dB below its baseline.

Procedure

Participants began by completing the same headphone test used in Experiments 1–3 (Woods et al., 2017). They were then given the main task instructions. Each trial featured the same relative tempo judgment task from Experiments 1–3, with the exception that the tapping manipulation was removed. Each participant completed three practice trials, as well as one repetition of each of the 180 unique combinations of timbre, pitch height, and stimulus tempo. Among the three presentations of each timbre × pitch height × tempo range pairing, one trial was played at each of the three loudness levels. Trials were arranged into six blocks of 30. All trials within a block used the same timbre, and the timbre alternated between blocks in an ABABAB pattern. We randomized which timbre was presented in the first block between participants, and we randomized pitch heights and stimulus tempos using the same constraints as were used in Experiments 1–3. The practice trials matched the timbre used in the first block, and their pitch heights and tempos were identical to those used in the practice trials for Experiments 1–3.

Response Scoring

Response scoring was identical to Experiments 1–3, with the exception that we reduced the Cook's distance threshold for excluding trials as outliers from 4/90 to 4/180 due to the trial count doubling in Experiment 4. A total of 709 trials (5.12%) were excluded as outliers via this metric.

Data Analysis

We applied the same general process from Experiments 1–3, in which we first analyzed all coefficients of our regression model together via a multivariate test, and then used univariate post hoc testing as necessary with Holm–Bonferroni correction. We tested the subject-level linear models of tempo ratings (Equation 2.3.2) against ground-truth relative tempo (Equation 2.3.1) via identical methods to Experiments 1–3. We also analyzed the effect of loudness on illusory tempo via the same methods as before, with the exception that no Hyunh-Feldt correction was applied, as there were no violations of sphericity.

To analyze the main effect of pitch height, we fit subject-level, second-order polynomial regression models to predict illusory tempo shift as a function of pitch height. We then tested the linear and quadratic slopes against zero via a one-sample Hotelling's T^2 test. To analyze the main effect of timbre, as well as whether the effect of pitch differed between timbres, we fit separate second-order polynomial regression models to each participant's data from each timbre. Then we compared the linear and quadratic slopes, as well as the intercepts, between timbre conditions, using a dependent-samples Hotelling's T^2 test. Significant differences in the intercepts can be understood as a main effect of timbre, whereas significant differences in the slopes

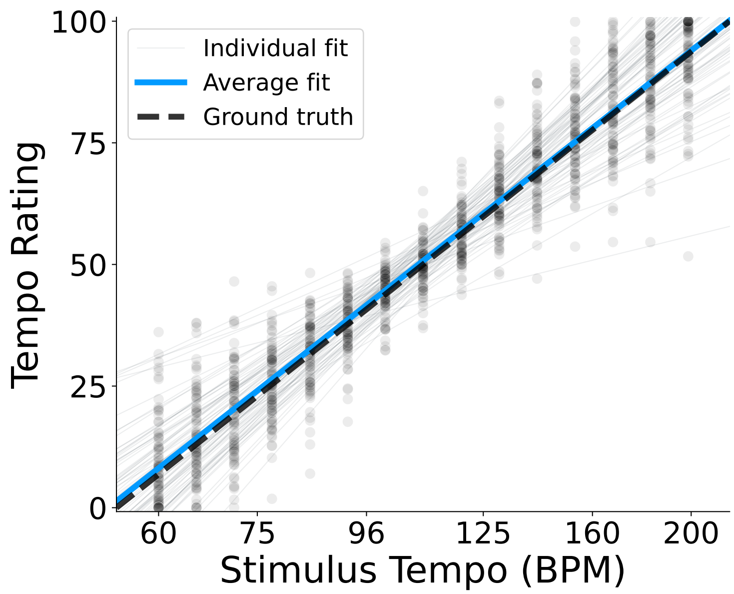

Figure 2.7: Points show participants' average ratings for each stimulus tempo in Experiment 4. Ground-truth ratings (Equation 2.3.1) are indicated by the dashed line. Regression lines for individual participants (Equation 2.3.2) appear as thin gray lines, and the average of all regression lines appears in blue.

can be understood as a significant interaction between pitch height and timbre.

2.4.2 Results

Subjective Ratings versus Ground Truth

Figure 2.7 illustrates raw tempo ratings as compared to the theoretical ground-truth ratings. The average of all subject-level regression models ($\beta_0 = 50.74$, $\beta_1 = 49.41$) did not differ significantly from the ground-truth (Equation 2.3.1), $F(2, 75) = 2.13$, $p = .126$, $\eta_p^2 = .054$. Participants were highly accurate on average at rating relative tempo.

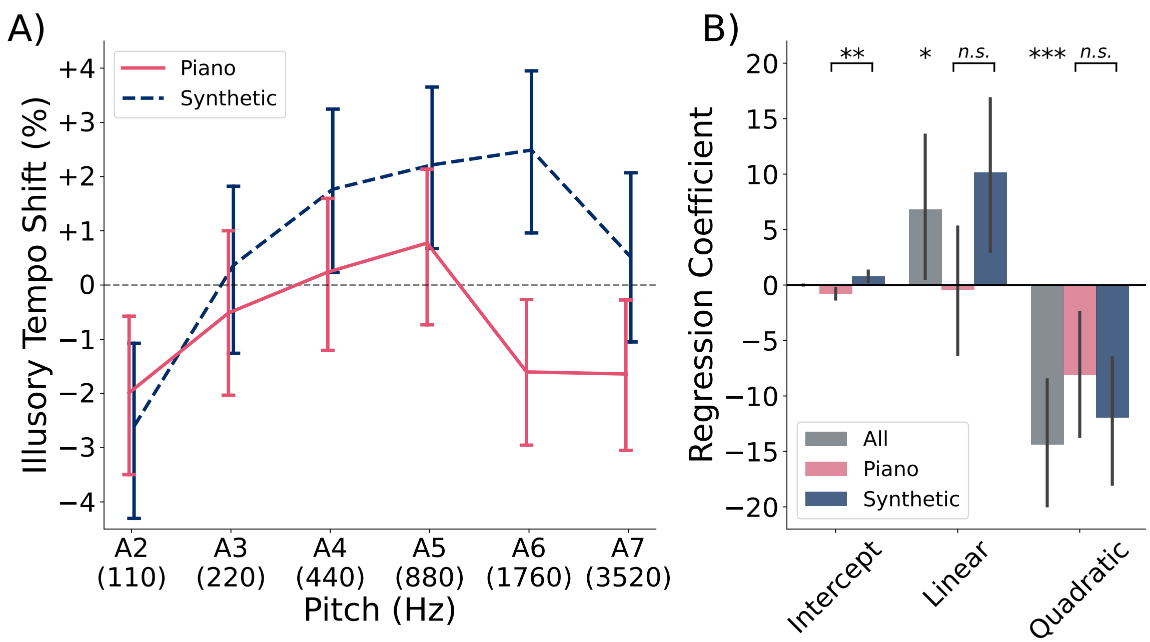

Figure 2.8: **A)** Average illusory tempo shifts across participants for repeating tones of each pitch height and timbre in Experiment 4. Data from piano-tone trials are shown in pink, and data from synthetic-tone trials are shown in blue. **B)** Average intercepts and slopes of subject-level polynomial regression models, predicting illusory tempo shift as a function of pitch height. Grey bars indicate coefficients for models fit to all of a participant's trials, pooling across timbres. Pink and blue bars indicate coefficients for models fit only to a participant's trials of one specific timbre. Error bars in both panels indicate 95% confidence intervals. The use of controlled, synthetic tones did not attenuate the quadratic component of the illusory tempo effect.

Pitch and Timbre

Our primary question of interest in Experiment 4 was whether synthetic tones would eliminate the negative quadratic effect of pitch height on relative tempo judgments. Such a result would suggest that the decrease in tempo ratings at extremely high octaves in Experiments 1–3 may have been due to low-frequency mechanical noise contaminating the piano recordings. Figure 2.8A shows the average illusory tempo scores for repeating tones of each pitch height, separated by timbre. A clear inverted U-shaped pattern emerged in the perceived tempo of both timbres. The corresponding polynomial regression coefficients are shown in Figure 2.8B. Collapsing across timbres, we observed a significant main effect of pitch height on perceived tempo, $F(2, 75) = 14.74$, $p < .001$, $\eta_p^2 = .282$, which was characterized by both a significant positive linear, $t(76) = 2.12$, $p = .037$, $d = 0.242$, and negative quadratic slope, $t(76) = -4.87$, $p < .001$, $d = 0.555$.

We also observed a significant difference in illusory tempo shift between timbres, $F(3, 76) = 5.09$, $p = .003$, $\eta_p^2 = .282$; however, this difference was characterized primarily by synthetic tones producing a significantly higher intercept than that of piano tones (i.e., a main effect of timbre), $t(76) = 3.38$, $p = .001$, $d = 0.720$, and a slightly more positive linear slope, $t(76) = 2.20$, $p = .031$, $d = 0.361$ (non-significant after correction for multiple comparisons). Rather than eliminate the quadratic effect of pitch height, the synthetic tones showed a more negative average quadratic slope than the piano tones, though not significantly so, $t(76) = -1.00$, $p = .320$, $d = 0.153$. These results confirm that the inverted U-shaped effect of pitch height on perceived tempo is not specific to the piano timbre used in Experiments 1–3.

Loudness Control

We again included a loudness control to assess whether subtle loudness differences between tones may have affected tempo ratings, and found no significant effect of loudness on perceived tempo, $F(2, 152) = 0.57$, $p = .567$, $\eta_p^2 = .007$. Mean illusory tempo shifts for tones played 3 dB below baseline, at baseline, and 3 dB above baseline were −0.13% ($SD = 2.96$), +0.36% ($SD = 3.20$), and −0.23% ($SD = 2.72$), respectively.

2.4.3 Discussion

In Experiment 4, we evaluated whether the inverted U-shaped effect of pitch height on tempo judgments was specific to the piano tones used in Experiments 1–3. Specifically, we considered whether the low salience of extremely high pitches (Terhardt et al., 1982) caused participants to focus instead on the low-frequency noise content underlying the highest piano tones. In Experiment 4, we generated synthetic tones without this low-frequency content and compared illusory tempo shifts for tones of both timbres. We found that synthetic tones were perceived as significantly faster than piano tones, but otherwise produced a similar inverted U-shaped effect (see Figure 2.8). These results suggest that the non-monotonic shape of the illusory tempo effect is not an artifact of low-frequency content in piano tones, and that the effect may generalize across timbres.

Our results do not rule out the possibility of a direct effect of pitch salience on perceived tempo; they only rule out a confounding effect in which low-frequency sounds from the mechanics of a piano are relatively salient when paired with extremely high-pitched tones. It remains noteworthy that the maximum of our illusory tempo curve falls near the 700 Hz peak of the pitch salience curve of Terhardt et al. (1982). Whether this similarity is a coincidence or evidence for an effect of salience on perceived tempo remains in question. It has been hypothesized that the energy expended in processing a stimulus biases perceived timing (Eagleman, 2008; Eagleman & Pariyadath, 2009), which makes it plausible that pitch salience might influence tempo perception. However, the results of Experiment 5 were inconsistent with this salience account, so we do not address it further.

2.5 Experiment 5

In Experiment 4, the removal of low-frequency noise from high-pitched tones failed to eliminate the negative quadratic effect of pitch on perceived tempo. The origin of this effect thus remains in question. Another possibility is that the quadratic effect is simply a midpoint effect, whereby people perceive the middlemost tones to be fastest and the most extreme tones to be slowest. We therefore conducted a fifth experiment in which we divided the pitch range from Experiments 1–4 in half, such that some participants heard only tones between A2 (110 Hz) and D♯5 (622.3 Hz), and others heard tones only between D♯5 and A7 (3520 Hz). We subdivided each register into six tones in order to match the six-tone structure of Experiments 1–4. Therefore, if the quadratic effect of pitch represents a midpoint effect, we should observe inverted U-shaped effects in both the lower and upper registers. In contrast, if the quadratic effect is an effect of absolute pitch or pitch salience, we should observe a positive linear effect of pitch in the lower register and a negative linear effect in the upper register.

2.5.1 Methods

Participants

Seventy-six undergraduate students (11 male, 64 female, 1 non-binary) from McMaster University completed the experiment for course credit. Ages ranged from 17–41 years ($M = 19.1$, $SD = 3.0$). Participants were randomly assigned to hear a register of lower octaves ($N = 37$; 33 female) or a register of upper octaves ($N = 39$; 31 female). All participants completed the task online between July and October 2021 due to the COVID-19 pandemic. Three additional participants completed the experiment but were excluded due to low correlation ($r < .5$) between their responses and the true relative tempo (see Response Scoring).

Materials

The metronome was identical to that used in Experiments 1–4. Tones were synthetic complex tones constructed in an identical manner to those in Experiment 4. All sound sequences were pre-generated and were played back as WAV files during the experiment.

Apparatus

We conducted the experiment online using the same apparatus as Experiments 1–4.

Design & Procedure

The procedure was identical to that of Experiment 4, with the exception that (1) tones in all six blocks were synthetic and (2) we added a between-subjects manipulation of register. The experiment therefore had a mixed 2 (Register) × 6 (Pitch Height) × 15

(Stimulus Tempo) design. Participants assigned to the lower register heard the six pitches A2 (110 Hz), D♯3 (155.6 Hz), A3 (220 Hz), D♯4 (311.1 Hz), A4 (440 Hz), and D♯5 (622.3 Hz), whereas those assigned to the upper register heard the six pitches D♯5 (622.3 Hz), A5 (880 Hz), D♯6 (1244.5 Hz), A6 (1760 Hz), D♯7 (2489.0 Hz), and A7 (3520 Hz).

Response Scoring

Response scoring was identical to Experiment 4. A total of 664 trials (4.85%) were excluded as outliers.

Data Analysis

We followed a similar procedure to our prior experiments, in which we first analyzed regression model coefficients together via multivariate tests, then conducted univariate post hoc testing with Holm–Bonferroni correction as necessary. We tested the subject-level linear models of raw tempo ratings, as well as the effect of loudness on illusory tempo shift via the same methods as Experiment 4.

To analyze the effect of pitch height on perceived tempo, we fit subject-level, second-order polynomial regression models to predict illusory tempo shift as a function of pitch height. To assess the main effect of pitch, we tested the linear and quadratic slopes from participants in each register condition against zero via one-sample Hotelling's T^2 tests for each register. To determine whether the effect of pitch differed between registers, we compared the linear and quadratic slopes between registers using an independent-samples Hotelling's T^2 test.

As average residual tempo and illusory tempo shift are always equal to zero at the

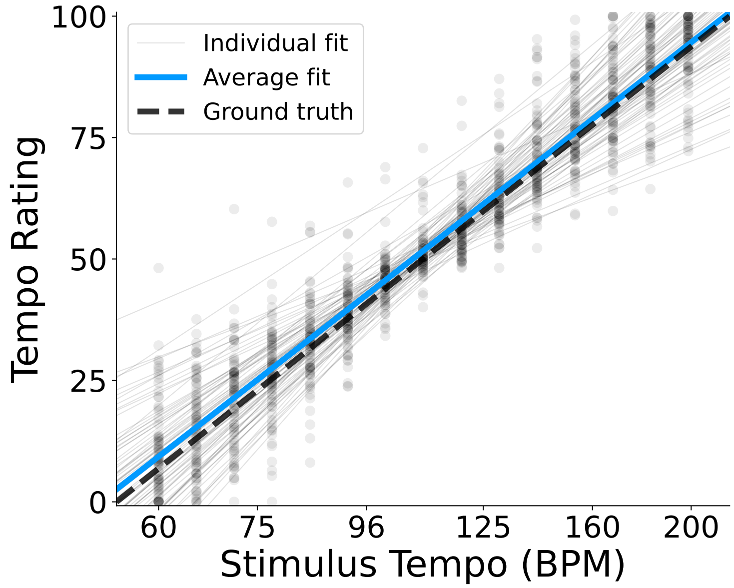

Figure 2.9: Points show participants' average ratings for each stimulus tempo in Experiment 5. Ground-truth ratings (Equation 2.3.1) are indicated by the dashed line. Regression lines for individual participants (Equation 2.3.2) appear as thin gray lines, and the average of all regression lines appears in blue.

subject level, the main effect of register on perceived tempo cannot be assessed by comparing the intercepts of the illusory tempo models of participants in each group. Instead, we compared the intercepts of the raw tempo rating models (Equation 2.3.2) between conditions, using an independent-samples t-test. This test captures the main effect of register by determining whether average raw tempo ratings were higher in one register than another.

2.5.2 Results

Subjective Ratings versus Ground Truth

Figure 2.9 illustrates raw tempo ratings from Experiment 5 as compared with the theoretical ground-truth ratings. The average linear fit ($\beta_0 = 51.66$, $\beta_1 = 49.10$)

differed significantly from the ground truth, $F(2, 74) = 6.91$, $p = .002$, $\eta_p^2 = .157$. Similar to Experiments 1–3, the intercept was significantly greater than the ground truth, $t(75) = 3.72$, $p < .001$, $d = 0.427$, $CI = [50.77, 52.55]$, while the slope did not significantly differ, $t(75) = -0.71$, $p = .479$, $d = 0.082$, $CI = [46.57, 51.62]$. Additionally, the average intercept among participants in the upper register ($M = 52.16$) was slightly higher than among those in the lower register ($M = 51.13$), but not significantly so, $t(74) = -1.15$, $p = .253$, $d = .264$. Participants were generally accurate at judging differences in relative tempo, but tended to slightly overestimate relative tempo overall, regardless of which register they heard.

Pitch and Register

Figure 2.10A illustrates illusory tempo shifts from Experiment 5 as a function of pitch height within each register. Contrary to the inverted U-shaped effect observed in Experiments 1–4, the data show a simple higher-equals-faster bias regardless of whether participants heard tones in the lower register only (A2–D♯5; 110–622.3 Hz) or the upper register only (D♯5–A7; 622.3–3520 Hz). Multivariate analysis of the regression coefficients (Figure 2.10) revealed significant main effects of pitch height in both the lower register, $F(2, 35) = 13.41$, $p < .001$, $\eta_p^2 = 0.434$, and upper register, $F(2, 37) = 4.99$, $p = .012$, $\eta_p^2 = 0.212$. In the lower register, post hoc univariate testing revealed a strong, positive linear effect of pitch on perceived tempo, $t(36) = 5.21$, $p < .001$, $d = 0.856$, with a nonsignificant quadratic effect, $t(36) = -1.91$, $p = .064$, $d = .315$. In the upper register, we observed a significant positive linear effect of pitch with medium effect size, $t(38) = 3.17$, $p = .003$, $d = 0.508$, and a nonsignificant quadratic effect of pitch, $t(38) = 0.72$, $p = .477$, $d = 0.115$. The

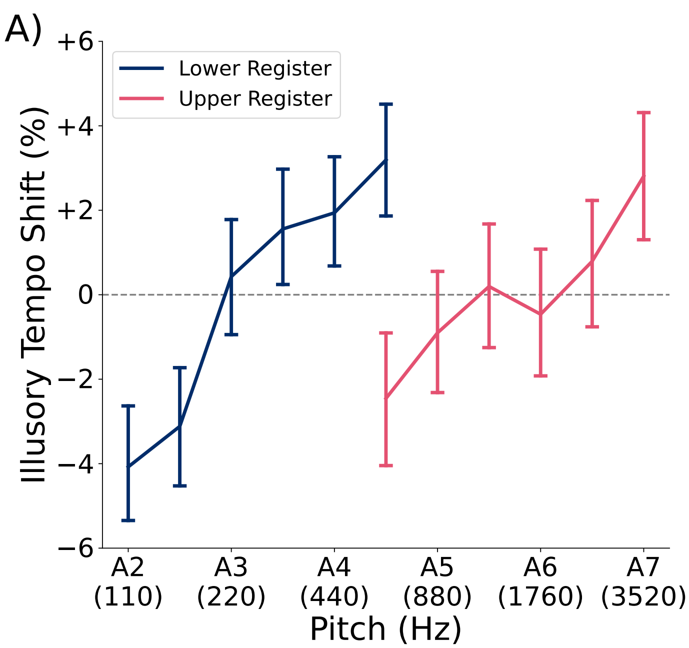

Figure 2.10: **A)** Line plots indicate illusory tempo shifts across pitch heights in Experiment 5, averaged across participants who heard repeating tones from the lower register only (blue) versus the upper register only (pink). **B)** Slope coefficients of subject-level polynomial regression models predicting illusory tempo shift as a function of pitch height. Blue and pink bars show the average coefficients for participants assigned to the lower and upper registers, respectively. Error bars in both panels indicate 95% confidence intervals. Participants rated higher pitches as faster in a positive linear pattern, regardless of the register of stimuli they listened to.

polynomial regression slopes did not differ significantly between participants assigned to the lower and upper registers (Figure 2.10B), $F(2, 73) = 2.62$, $p = .080$, $\eta_p^2 = .067$, suggesting that pitch height influenced perceived tempo similarly in both registers. In both groups, higher pitches were perceived as faster in a positive linear pattern.

Loudness Control

We included the same loudness control as in Experiments 1–4 to assess whether subtle loudness differences between tones may have affected tempo ratings, and we again found no significant effect of loudness on perceived tempo, $F(2, 150) = 0.45$, $p = .639$, $\eta_p^2 = .006$. Mean illusory tempo shifts for tones played 3 dB below baseline, at baseline, and 3 dB above baseline were +0.19% ($SD = 2.07$), −0.02% ($SD = 1.72$), and −0.17% ($SD = 4.04$), respectively.

2.5.3 Discussion

In Experiment 5, we observed a linear higher-equals-faster bias regardless of whether participants heard tones from lower octaves or upper octaves only. These results are inconsistent with both the hypothesis that the inverted U-shaped effect of pitch depends on a midpoint effect, as well as the hypothesis that it depends on absolute pitch or pitch salience. Rather, our results suggest that the illusory effects of pitch on perceived tempo depend on relative pitch within a context. The finding that register did not significantly affect average raw tempo ratings further supports this importance of relative pitch over absolute pitch.

It also appears that the effect may depend on the range of pitches included in a context. When participants in Experiments 1–4 heard tones spanning a five-octave range, they demonstrated an inverted U-shaped effect of pitch, in which perceived tempo increased with pitch for two to three octaves before reaching a plateau and ultimately declining again. In contrast, when participants in Experiment 5 heard tones spanning only two and a half octaves, perceived tempo simply increased linearly with pitch across the entire range, regardless of whether that context was low or high in absolute pitch. We address potential explanations for this context dependence in the General Discussion, below.

2.6 General Discussion

We conducted five experiments to better understand the shape of the relation between pitch and perceived tempo across several octaves. Previous research has identified pitch-induced illusory tempo effects (Boltz, 2011), but only tested these effects within a relatively restricted range of fundamental frequencies (see Figure 2.1). Our goal was to map out the shape of these effects across the broader span of human hearing, and to evaluate the generalizability of the claim that higher pitches are perceived as faster. Our results instead suggest a context-dependent effect, in which perceived tempo increases with pitch until peaking between two and three octaves above the lowest pitches in the context. Above this point, the pitch-induced bias reverses direction, resulting in tones five octaves above the lowest pitch being perceived as slower than those two octaves above. In contexts spanning fewer than three octaves, we observed a monotonic higher-equals-faster relationship regardless of the absolute pitch of the context. Furthermore, we found that pitch-induced illusory tempo was consistent regardless of the true tempo, that it was only slightly attenuated via motor synchrony, and that it appeared with both piano tones and synthetic complex tones.

Our results raise the question of why the context's pitch range matters. One possibility is that the effects we observed are tied to human vocal ranges. Individuals typically have vocal ranges that span about two octaves (Kuhn, Wachhaus, Moore, & Pantle, 1979; Moore, 1991), which is similar to the range over which we observed a bias to rate higher-pitched tones as faster. There are individual and gender differences in the absolute pitch of human voices, so it may be that relative pitch within a speaker or singer's vocal range better predicts their vocal timing than does the absolute pitch of their voice. Analysis by Broze and Huron (2013) has similarly suggested that relative pitch in Western music better predicts timing than does absolute pitch. Perhaps, then, relative pitch within a context drives illusory tempo because it is informative as a tempo cue. Additionally, as relative pitch may be most meaningful within the range of an individual human voice or typical musical melody, the perceived increase in tempo alongside increasing pitch may primarily apply within a roughly two-octave range.

We are left, then, with the question of how pitch-induced illusory tempo arises, and at which stage of processing. One possibility is that auditory dimensions might interact due to the physics of the peripheral auditory system – similar to interactions between frequency and amplitude that have been observed in the cochlea in the form of frequency detuning (Large, 2010; Ruggero, 1992). Indeed, at least one simulation of auditory nerve and midbrain responses to periodic stimuli has demonstrated a U-shaped effect of stimulus frequency on the strength of neural synchronization to a beat (see Zuk, Carney, & Lalor, 2018, Supplementary Figure 3). However, the context sensitivity of illusory tempo, as observed in Experiment 5, would be difficult to reconcile with an origin in the dynamics of the auditory periphery.

Therefore, we believe it is most plausible that these pitch-time interactions depend on perceptual experience, as suggested by Boltz (1998), and that they are learned from real-world correlations between pitch and tempo. If these effects do derive from perceptual experience, it may be possible to neutralize or invert the illusion in the short term if, for example, people listened to a selection of music where pitch and tempo were negatively correlated. The developmental trajectory of pitch-time integration also remains an open question, and future research should attempt to measure whether illusory tempo effects are present in infants and young children. Research by Eitan and colleagues on imagined and physical motion while listening to music has at most found weak associations between pitch and movement speed in children as old as 11 (Eitan & Tubul, 2010; Kohn & Eitan, 2009, 2016), suggesting that higher-faster associations may develop later in childhood.

One limitation of our relative tempo judgment task is that it cannot determine whether differences in tempo ratings arise at the level of perception or decision-making. That is, we cannot determine whether participants rated some tones as faster than others because a perceptual bias caused them to hear the tones as faster, or because their decision-making strategy used pitch as a post hoc cue (e.g., "the tones were low, so they were probably slow"). Future work should test perceived tempo as a function of pitch in a task without a decision-making component. One approach would be to use a sensorimotor paradigm, such as the synchronization and continuation tapping tasks used by Boasson and Granot (2012, 2019) and Ammirante et al. (2011), respectively. If pitch-induced illusory tempo arises at the level of perception, participants' tapping rates should parallel the tempo ratings observed in the present study. However, if these biases arise at the level of decision-making, they should not influence synchrony. Given earlier findings that contour changes, pitch distances, and pitch direction do influence tapping behavior (Ammirante et al., 2011; Boasson & Granot, 2012, 2019), we might expect to find that pitch height also influences the tempo of synchronized motor behavior.

Outside of Experiment 4, we also did not attempt to disambiguate whether our effects derive from the pitch of a sound, versus its frequency content. The perceived pitch of a sound depends not only on its fundamental frequency but also its harmonics, and tones with missing fundamentals and harmonics can still be perceived as having the same pitch (Fletcher, 1924). Future research should test whether perceived tempo varies solely with perceived pitch, or whether identically-pitched tones that lack high- or low-frequency harmonics would also be perceived as differing in tempo. Disambiguating the effects of pitch and frequency may help us to identify which stage of perceptual processing gives rise to illusory tempo effects.

Understanding interactions between musical dimensions is an important step towards linking models of pitch and time perception, which have historically been developed independently (e.g., Large et al., 2015; Large, Kim, Flaig, Bharucha, & Krumhansl, 2016), despite evidence for the integration of these processes. We also wish to stress the importance of our findings for research design in auditory perception. Analyses that assume pitch and timing are perceptually independent ignore a growing body of evidence that one can influence the other. It has become standard procedure to control for differences in perceived loudness between pitches (Fletcher & Munson, 1933) when designing experiments. Perhaps pitch-time interactions should be considered, as well.

Chapter 3

Pitch biases sensorimotor synchronization to auditory rhythms

3.1 Preface

As noted at the beginning of Chapter 2, our subjective tempo rating task was originally intended to be paired with a measure of sensorimotor timing, in a two-session format. By doing so, we hoped to gain insight into the stage of processing from which illusory tempo originates. For example, an effect that arises during decision-making might impact subjective ratings of tempo without altering sensorimotor timing, whereas an effect that arises on-line during perceptual processing would likely impact both.

Although the closure of in-person testing between March 2020 and February 2022 prevented us from collecting sensorimotor data from the same participants as our subjective rating experiments reported in Chapter 2, we ultimately did conduct two finger-tapping experiments between April 2022 and October 2023. These experiments were designed to measure the perceived tempo of the same pitches participants rated in Chapter 2, but this time in the absence of explicit decision-making. Chapter 3 outlines our findings and interprets them in light of the data reported in Chapter 2.

This chapter has been released as a preprint on *PsyArXiv* at `https://osf.io/preprints/psyarxiv/fbmw5`, and was submitted on July 19, 2024 as a manuscript for a special collection in *Scientific Reports* entitled "Auditory processing and perception." In this chapter, we reference the studies reported in Chapter 2 as Pazdera and Trainor (2024b) and Chapter 4 as Pazdera and Trainor (2023).

3.2 Introduction

In the field of music perception, there are a several dominant models of how humans synchronize their movements to rhythmic sounds (Large et al., 2023; Palmer & Demos, 2022). *Rhythm* refers to a series of events that are predictably organized across time, and humans—as well as some non-human animals (Gámez et al., 2018; Patel et al., 2009a; Patel, Iversen, Bregman, & Schulz, 2009b)—are able to extract a sense of underlying *pulse* or *beat* from these predictable event structures. Some models envision this process of perceiving and synchronizing to the beat as an emergent property of neural dynamics, in which the auditory and motor systems resonate when driven by rhythmic stimulation (Large & Snyder, 2009; Large et al., 2010, 2015; Roman, Roman, Kim, & Large, 2023). Others conceptualize the brain as a Bayesian prediction engine that works to infer the structure underlying predictable patterns in time, in accordance with predictive coding theory (Cannon, 2021; Friston, 2010; Koelsch et al., 2019; Vuust & Witek, 2014). Still others propose biophysical pacemaker models that learn stimulus timing through error correction (Bose, Byrne, & Rinzel, 2019; Egger, Le, & Jazayeri, 2020).

Regardless of category, these established models track rhythms based on the phase, period, and amplitude of sound patterns, but not their pitch. However, there is a wealth of evidence to suggest that pitch affects the perceived tempo of a rhythm (Boltz, 2011; Collier & Hubbard, 1998; Pazdera & Trainor, 2024b), as do the size (Ammirante et al., 2011; Ammirante & Thompson, 2012; Boasson & Granot, 2012; Boltz, 1998) and direction of pitch changes (Boasson & Granot, 2019; Gordon & Ataucusi, 2021; Herrmann et al., 2013, 2014). In the present study, we sought to identify whether previously identified effects of pitch height on subjective tempo ratings (Pazdera & Trainor, 2024b) also influence the timing of sensorimotor synchrony. If they do, it would support the incorporation of pitch as a factor in models of rhythm perception, whether as an influence on neural dynamics or as an informative cue in a Bayesian prediction process.

3.2.1 Integration of Pitch and Timing

The idea that pitch and time perception are integrated dates back decades, originating as an extension of research on the *kappa* effect, in which observing spatial movement distorts time perception (Cohen, Hansel, & Sylvester, 1953; Cohen et al., 1954). This concept of integrality was formalized by Jones (1976), who proposed that the brain integrates pitch, time, and loudness into a shared representational space. She hypothesized that music is represented as a trajectory of movement through this multidimensional space, and lawful relations between movements along different dimensions guide musical expectancy. A similar idea underlies the auditory pitch-motion hypothesis of Henry and McAuley (2013), in which the movements of an auditory signal in frequency space guide the predicted timing of that signal. Boltz (2017) has further suggested that pitch and time share an integrated representation in memory, and our own lab has recently presented evidence that pitch and timing bidirectionally influence one another in auditory perception (Pazdera & Trainor, 2023).

In general, higher pitch has been associated with faster and earlier perceived timing (Boltz, 2011, 2017; Collier & Hubbard, 1998; Gordon & Ataucusi, 2021; Herrmann et al., 2013, 2014; Pazdera & Trainor, 2023). For example, Collier and Hubbard (1998) found that participants rated higher-pitched repeating tones and musical scales as faster and as speeding up more than those played in a lower octave. Similarly, Boltz (2011) found that participants tended to rate melodies as faster when played in a higher octave, compared to a lower one. Furthermore, they rated ascending melodies as faster than descending melodies. Herrmann et al. (2013, 2014) have since discovered neural correlates of this pitch-induced illusory tempo effect in both magnetoencephalography (MEG) and functional magnetic resonance imaging. Most relevant to models of rhythm perception is their discovery that patterns of ascending and descending pitch shift the phase of entrained neural activity in opposite directions relative to a rhythmic auditory stimulus, and improve neural tracking of tempo changes in the congruent direction (Herrmann et al., 2013).

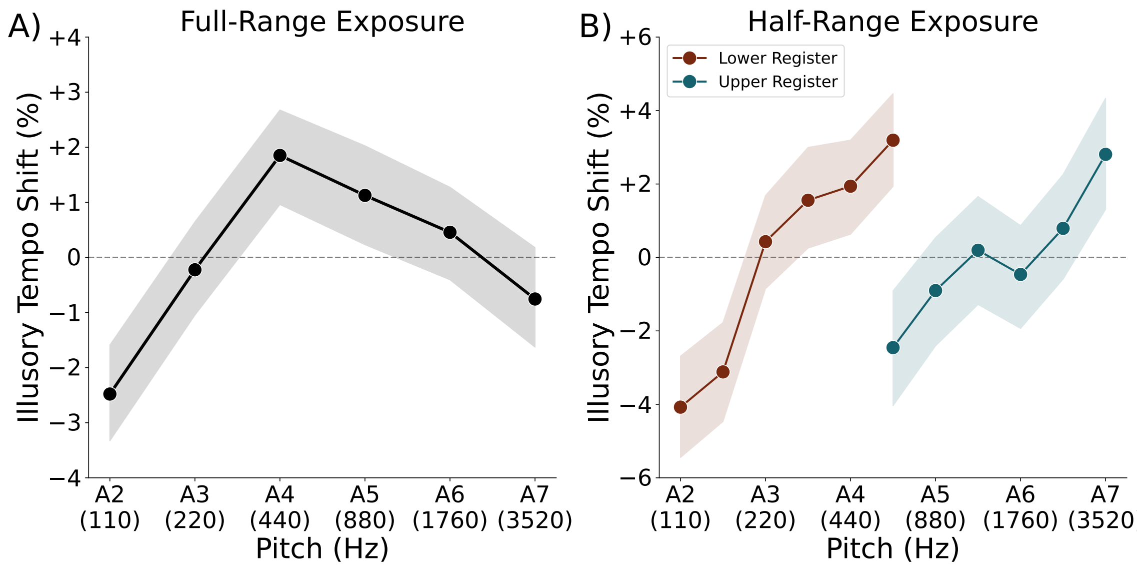

Figure 3.1: Subjective tempo ratings for isochronous repeating tones as a function of pitch, as observed when participants **A)** heard tones between 110 Hz and 3520 Hz (Pazdera & Trainor, 2024b, Experiments 1–3), or **B)** heard tones from only the lower or upper half of that range (Pazdera & Trainor, 2024b, Experiment 5). Shaded bands indicate 95% confidence intervals. Exposure to a five-octave range of frequencies produced an inverted U-shaped relation between pitch and perceived tempo, whereas exposure to a 2.5-octave range produced a bias to perceive higher pitches as faster regardless of absolute pitch.

3.2.2 Nonmonotonicity of Pitch-Induced Timing

We recently conducted a study testing the generalizability of the aforementioned findings by Collier and Hubbard (1998) and Boltz (2011) to a wider range of frequencies than previously tested (Pazdera & Trainor, 2024b). Whereas the former compared the perceived timing of the pitches C4 (261.6 Hz), C5 (523.2 Hz), and C6 (1046.4 Hz) and the latter tested melodies starting on C3 (130.8 Hz) or C5 (523.2 Hz), we tested tempo perception across pitches as low as A2 (110 Hz; close to the average male speaking voice) to as high as A7 (3520 Hz; the fourth-highest note on a piano). In our study, we asked participants to rate how fast repeating tones were compared to a metronome. The metronome had the same tempo on every trial, acting as a standard reference, whereas the repeating tone varied in pitch and tempo across trials. Contrary to the generally accepted view that higher pitches are perceived as faster, we observed an inverted U-shaped relation between pitch and perceived tempo (Figure 3.1A). Between 110 Hz and 440 Hz, higher pitches were consistently perceived as faster; however, perceived tempo peaked somewhere between 440 Hz and 1760 Hz (varying across experiments), with 3520 Hz reliably perceived as slower than 440 Hz. These data suggest that the relation between pitch and perceived tempo is in fact nonmonotonic, and reverses direction at the upper octaves left untested by previous studies.

To address the possibility of a midpoint effect in which people perceive the middlemost pitches in any context as fastest, we next exposed one group of participants to pitches only between 110 Hz and 622.3 Hz and another group to pitches only between 622.3 Hz and 3520 Hz (Pazdera & Trainor, 2024b, Experiment 5). If people perceive the average pitch in a context as fastest, we expected participants in both registers to show inverted U-shaped illusory tempo effects. If illusory tempo instead depends on absolute pitch, we expected participants assigned to the lower register to rate higher pitches as faster and participants in the upper register to rate higher pitches as slower. Unexpectedly, participants in both registers rated higher pitches as faster, eliminating the U-shaped effect entirely (Figure 3.1B). We interpreted this finding as an indication that illusory tempo depends on the pitch range within a context. Within a context spanning 2–3 octaves, higher pitches were consistently perceived as faster; however, pitches several octaves above baseline were instead perceived as slower, perhaps tying the effect to implicit knowledge about the span of human vocal ranges (Kuhn et al., 1979).

3.2.3 The Present Study

One limitation of a subjective rating paradigm like those previously used by ourselves and others (e.g., Boltz, 2011; Collier & Hubbard, 1998; Pazdera & Trainor, 2024b) is that they cannot distinguish whether biases in people's ratings arise at the level of perception or decision making. That is, we cannot tell in Figure 1 whether participants rated certain tones as faster because they actually experienced them as faster in real-time while listening to them, or because they implicitly (or explicitly) used pitch as post hoc evidence for how fast the tones likely were. To address this limitation, we designed the present study to test the perceived tempo of the same pitches as Pazdera and Trainor (2024b) in the absence of an explicit decision-making task. If our previously-observed effects are preserved in the absence of decision-making, then it would provide evidence for a perceptual origin of pitch-induced illusory tempo.

Outside of subjective ratings, the most common measurement of human timing is through sensorimotor tasks, usually involving tapping (Repp, 2005; Repp & Su, 2013). Both synchronization (Boasson & Granot, 2012, 2019) and continuation (Ammirante & Thompson, 2010, 2012; Ammirante et al., 2011) tapping tasks have been used previously to study the effects of pitch distance, pitch direction, and contour changes on perceived tempo. We therefore adopted a synchronization–continuation tapping task for the present study, to determine whether pitch height also influences sensorimotor timing.

In a synchronization–continuation task, participants first synchronize their movements to a metronome or other pacing signal, and then try to continue moving at the same tempo even after the signal ends (e.g., Michon, 1967; L. T. Stevens, 1886; Wing & Kristofferson, 1973) or switches to the participant's control (e.g., Ammirante et al., 2011; Flach, 2005). If the pitch of the pacing signal biases the perception of its tempo, then participants should synchronize to different pitches as if they were played at different tempos. For example, participants might tap at an earlier phase when synchronizing to pitches they perceive as speeding up (Boasson & Granot, 2012, 2019), and may tap faster when asked to continue moving at the tempo of the pacing signal (Ammirante & Thompson, 2010, 2012; Ammirante et al., 2011). Therefore, if pitch biases timing at the level of perception, we should observe a greater negative mean asynchrony (the extent to which tapping anticipates the stimulus) and shorter inter-tap intervals for the same pitches that participants rated as fastest in the Pazdera and Trainor (2024b) experiments.

In the present study, we conducted two experiments to test whether the same effects of pitch on subjective tempo ratings would carry over to sensorimotor timing in a synchronization–continuation task. In Experiment 1, we attempted to replicate the U-shaped effect from Pazdera and Trainor (2024b) Experiments 1–3 (Figure 3.1A) by exposing participants to pitches spanning a five-octave range from A2 (110 Hz) to A7 (3520 Hz). In Experiment 2, we attempted to replicate the simple higher–faster bias from Pazdera and Trainor (2024b) Experiment 5 (Figure 3.1B) by exposing participants to tones from only the lower half or upper half of that range.

3.3 Experiment 1

In Experiment 1, we tested whether the nonmonotonic relation between pitch and perceived timing (Pazdera & Trainor, 2024b) persists in the absence of an explicit decision-making task. Specifically, we asked participants to synchronize with eight isochronous tones that varied in pitch across trials, ranging from A2 (110 Hz) to A7 (3520 Hz), and then to continue tapping at the same tempo for 16 additional beats. If pitch affects subjective tempo at the level of perception, we should observe a similar pattern of results to those in the subjective rating task by Pazdera and Trainor (2024b). However, if pitch affects subjective tempo at a later stage of processing—such as during decision-making—we should not observe the same effect of pitch on sensorimotor timing.

3.3.1 Methods

Participants

Data for Experiment 1 were collected between April 2022 and January 2023, under special COVID-19 safety protocols, including mask requirements for all participants and experimenters. Thirty-eight undergraduate students (31 female, 6 male, 1 unreported) from McMaster University completed the experiment for course credit. Ages ranged from 17-30 (M = 18.7, SD = 2.2). An additional three participants completed the experiment, but were excluded based on performance criteria (see Data Analysis for details).

Data Availability

We have made all data, code, and stimuli from both experiments publicly available on the Open Science Framework at `https://osf.io/7bptg/`, as well as on GitHub at `https://github.com/jpazdera/IllusoryMotor`.

Materials

All text used in the experiment was displayed in white, 72-point Arial font on a black background. All auditory stimuli were 250 ms complex tones, which we generated in Python by summing sine waves for the fundamental frequency and the first two overtones. We summed the sine waves with random phase, and reduced the amplitude of overtones by 6 dB per octave. We then applied a percussive amplitude envelope (Schutz & Vaisberg, 2012) by adding a 10 ms linear rise followed by a 240 ms exponential fade. We applied an additional linear fade to the final 10 ms of the tone so that the amplitude ended at zero. Finally, we normalized the loudness of all tones to 75 dBA using Audacity. For Arduino compatibility, tones were saved as WAV files with a sampling rate of 22 kHz instead of the standard 44.1 kHz (see Apparatus). To account for the reduced sampling rate, we ensured that none of the frequencies used in our tone design (the highest being 10560 Hz, the second overtone of A7) exceeded the Nyquist frequency of our audio (11025 Hz).

Apparatus

We conducted audiometry using a Grason-Stadler GSI-61 Clinical Audiometer. During the main experimental task, an Arduino Uno running custom C++ code controlled auditory stimulus presentation and tap detection. The Arduino used two attachments: an Adafruit Wave Shield for audio presentation and an Ohmite FSR01CE force sensitive resistor (FSR) as a tapping pad (see Schultz & van Vugt, 2016). Stimuli were presented over Senheiser HD280 Pro headphones plugged into the Wave Shield, and participants tapped on the FSR. The Arduino communicated over USB with a Windows 7 computer running a Python (version 3.6) program that used the PsychoPy library (Peirce et al., 2019) to control the trial order and to log all data collected. The computer sent the Arduino instructions regarding which experimental condition to use on each trial, and the Arduino returned a timestamp for each stimulus onset, tap onset, and tap release. All text was displayed on a 19-inch Dell 1908FP monitor with a resolution of 1280 $\times$ 1024 and a frame rate of 60 Hz.

Design

Experiment 1 followed a 6 (Pitch Height) $\times$ 2 (Interonset Interval; IOI) within-subjects design. The pitch of the stimulus varied randomly between trials, and was one of A2 (110 Hz), A3 (220 Hz), A4 (440 Hz), A5 (880 Hz), A6 (1760 Hz), or A7 (3520 Hz) on each trial. The IOI of the synchronization tones also varied randomly between trials, and was either 600 ms (100 BPM) or 400 ms (150 BPM).

Procedure

Prior to the main synchronization–continuation task, we collected an audiogram from each participant in a sound-attenuated booth adjacent to the main testing room. Hearing thresholds were tested for pure tones at the frequencies 125 Hz, 250 Hz, 500 Hz, 1000 Hz, 2000 Hz, 4000 Hz, 8000 Hz. We first tested frequencies in ascending order from 1000 Hz to 8000 Hz, then tested in descending order from 1000 Hz to 125 Hz. At each frequency, we measured hearing threshold via a staircase procedure that began at +30 dB HL, reduced by 10 dB for every correct detection (to a minimum of -10 dB HL), and increased by 5 dB for every miss. We averaged the two measurements taken at 1000 Hz to obtain a single threshold value for this frequency.

The participant then returned to the main testing room, and we measured their spontaneous motor tempo by asking them to tap on the FSR with their index finger "at the rate that feels most natural and comfortable" to them. A fixation cross appeared on the screen to indicate when they should begin tapping, and the measurement ended after 30 taps.

We then introduced the participant to the synchronization–continuation task. We informed them that they would hear several tones playing at a steady pace on each trial, and that after several beats they would gain control of the tones so that their tapping would cause the tones to play. We emphasized to the participant that their goals were to first synchronize their taps so that their finger "lands exactly when the next tone will begin," and then to "keep the tones playing at the same steady pace throughout the trial by continuing to tap at the same rate" even after gaining control of the tones.

Each trial consisted of 24 repetitions of a tone, which varied in pitch across trial.

A fixation cross preceded the first tone of the trial by a uniformly jittered 1000 to 1500 ms. The first eight tones of the trial (the "synchronization tones") then played at an isochronous IOI of either 600 ms or 400 ms. The remaining 16 tones (the "continuation tones") were instead initiated by the participant's tapping. Continuation tones played 20 ms after the participants' taps, a delay which we introduced in order to help stabilize the transition from synchronization to continuation (Flach, 2005). If the participant tapped while the previous tone was still playing, no new tone would be generated until the participant lifted their finger off the FSR and tapped again. Immediately after the final continuation tone finished playing, the fixation cross disappeared and participants were shown the standard deviation (in milliseconds) of their continuation-phase inter-tap intervals alongside instructions to keep this score as low as possible by tapping steadily (Ammirante et al., 2011). This score remained onscreen for 2 s, after which a blank screen was displayed for 2 s before the next trial began.

Participants completed six practice trials and 120 experimental trials of the synchronization-continuation task. All practice trials used an IOI of 500 ms, and each used a different one of the six pitch height conditions, randomly ordered. The experimental trials were divided into five blocks of 24. Each block contained two trials of each pitch height and IOI combination, randomly ordered. Self-paced breaks were intended to be administered between blocks, but due to a software error were instead administered after the 12th, 24th, 36th, and 48th trials.

Data Analysis

Tap Debouncing To pre-process our tapping data, we first removed any falsely-detected taps through a debouncing procedure. The force sensitive resistor sometimes erroneously detected multiple onsets from a single tap if the participant's finger bounced upon landing, or if the tap pressure fluctuated above and below the minimum detection threshold during initial contact. To avoid analyzing these false taps and releases, we programmed the Arduino to automatically ignore any below-threshold pressure reading within 20 ms of a tap onset, treating the participant's finger as remaining on the FSR. Similar pressure fluctuations could also occur as the participant lifted their finger off of the resistor; therefore, the Arduino also ignored any above-threshold pressure reading within 80 ms of the previous tap's release, as if the participant's finger remained off the FSR. A review of our tap duration data suggested that some additional false releases occurred up to 30 ms after tap onset, and so were not automatically rejected by the Arduino. Any time the participant's finger remained on the FSR for longer than the 80 ms debounce period, a false tap onset was detected exactly 80 ms after these false releases. We therefore excluded from further analysis any tap that began 80 ms after a tap whose recorded duration was 30 ms or less.

After applying these debouncing methods, we separated taps into synchronization taps and continuation taps. We categorized any tap occurring between the third and ninth tone as a synchronization tap (including the tap that triggered the ninth tone), and any tap occurring after the ninth tone as a continuation tap. We excluded taps before the third tone, as a review of the data suggested that synchronization tended to stabilize during the third inter-stimulus interval.

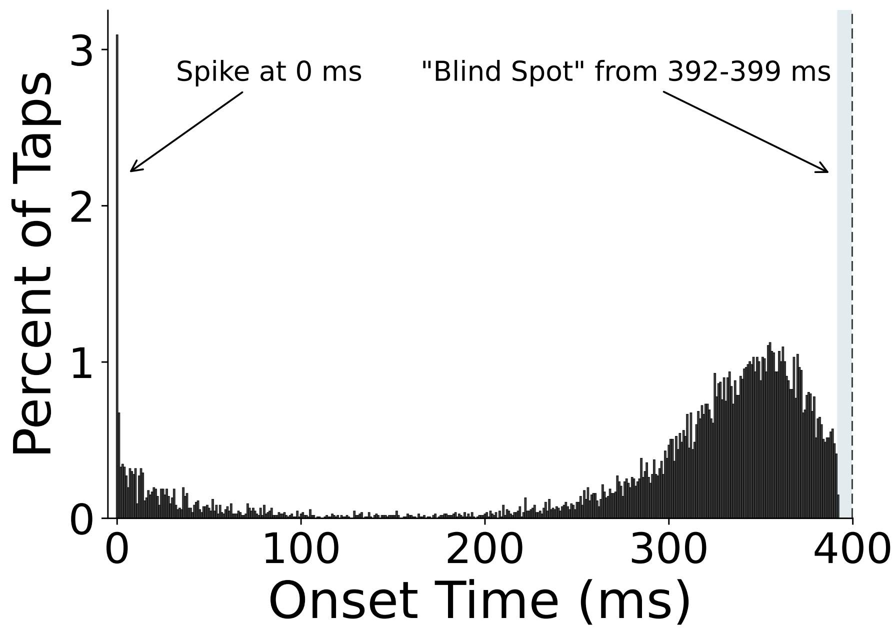

Figure 3.2: Distribution of synchronization taps recorded across all trials with an interonset interval of 400 ms. Onset time indicates the number of milliseconds that a tap was detected after the start of the most recent synchronization tone. Recorded taps across most onset times follow a circular-normal distribution; however, the inability of the Arduino to record taps while initializing the next stimulus results in an 8 ms gap (shaded region) leading up to the next tone (dashed line). All taps that begin during this "blind spot" are instead recorded immediately after the next tone begins, resulting in the spike at 0 ms post-stimulus. In Experiment 2, we reduced the pre-stimulus blind spot to 2 ms (see Experiment 2 Apparatus).

Synchronization Tap Processing For each synchronization tap, we identified the pair of tones between which it occurred, and calculated the fraction of the stimulus IOI that had elapsed before the tap occurred. We then converted this fraction to a phase value relative to the preceding tone, from 0 to 2π. If the tap that generated the first continuation tone occurred later than one IOI after the eighth synchronization tone (i.e., a relative phase greater than 2π), we excluded it from further analysis.

Because the Arduino can only perform one process at a time, it can only read from the FSR in between other operations. As such, longer operations can create "blind spots" in tap detection. In particular, we identified two operations lasting longer than 1 ms: loading an audio file from the Wave Shield's memory card took approximately 6 ms, and initializing audio playback took approximately 2 ms. In Experiment 1, file loading occurred immediately before playback; therefore, whenever the participant tapped within the 8 ms prior to a tone onset, the Arduino would instead detect the tap immediately after tone onset, producing an apparent relative phase of 0. To illustrate, Figure 3.2 shows the distribution of synchronization tap onset times relative to the preceding tone across all trials with an IOI of 400 ms. A review of this distribution suggests that the majority of taps recorded as occurring at the same millisecond as a tone were actually taps initiated during the preceding blind spot. We therefore excluded from further analysis all synchronization taps with an apparent relative phase of 0.

We next converted each tap's phase within its inter-stimulus interval into an asynchrony relative to the tone at which it was most likely targeted. Figure 3.3 illustrates this conversion process in one example participant. For each participant, i, we first calculated the circular mean of their tapping phases, μ_i, as well as the anti-phase of their average tap, $\mu_i - \pi$ (the dashed and dotted lines in Figure 3.3A, respectively). Our conversion method next relies on two assumptions: 1) taps at the mean phase were made in anticipation of the subsequent tone (as participants were instructed), and 2) each person's tap timing is symmetrically distributed around the mean. Under these two assumptions, any taps with a relative phase in the range $[\mu_i - \pi, 2\pi)$ (i.e., taps to the right of the dotted line in Figure 3.3A) were most likely targeted at the subsequent tone, and any taps with a relative phase in the range $(0, \mu_i - \pi)$

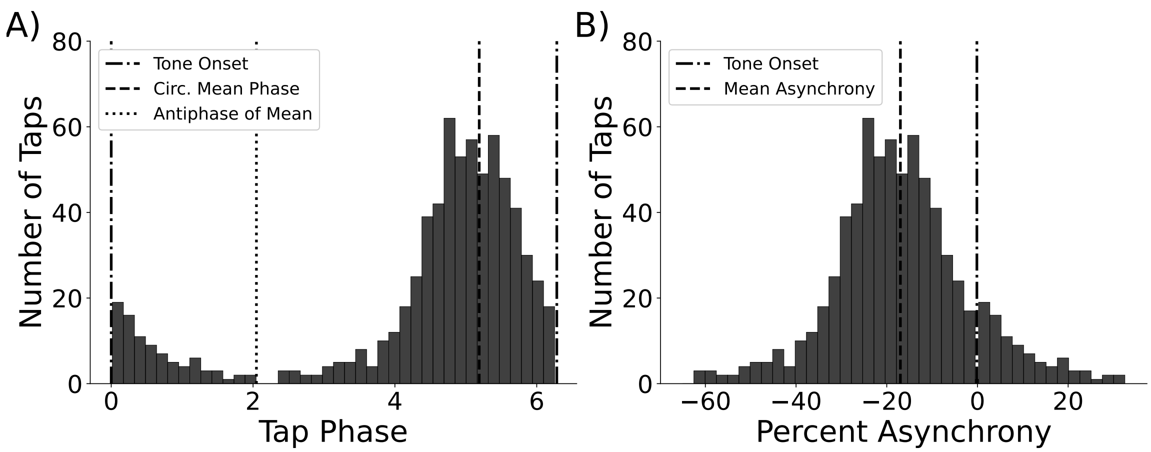

Figure 3.3: **A)** The distribution of synchronization tap phases, relative to the preceding tone onset, from one example participant. The dashed line denotes their circular mean tapping phase, while the dotted line indicates the antiphase of their tapping. We treated all tap times that were closer to the preceding stimulus than to the participant's mean tapping phase (i.e., all taps to the left of the antiphase line) as positive asynchronies targeted at the preceding tone. We treated all other tap times as negative asynchronies made in anticipation of the next tone. **B)** The resulting asynchrony values, expressed as a percent of the interonset interval of the stimulus.

(those to the left of the dotted line) were most likely targeted at the preceding tone. To ensure that the first assumption held, we excluded one participant whose mean phase fell closer to the preceding tone than to the subsequent tone and two additional participants with a mean resultant vector length less than 0.45 (indicating poor synchronization). Having aligned taps with the tones they most likely targeted, we concluded by calculating the percent of an interonset interval that each tap occurred before or after its target. This produced a percent asynchrony for each tap (Figure 3.3B), where a negative asynchrony indicates that the tap preceded its target tone and a positive asynchrony indicates the tap occurred after its target.

Continuation Tap Processing For each continuation tap, we calculated the preceding inter-tap interval as the number of milliseconds elapsed since the onset of the previous tap. We then divided the inter-tap interval by the IOI of the synchronization tones on that trial to obtain a score we will refer to as the *relative inter-tap interval*. A relative inter-tap interval of 1 would indicate perfectly accurate continuation tap timing, with the inter-tap interval being equal to the IOI. Scores greater than 1 indicate that the participant tapped slower than the synchronization tones, whereas scores less than 1 indicate faster tapping. To eliminate intervals where the participant paused or tapped excessively quickly, we excluded from further analysis any continuation tap with a relative inter-tap interval greater than 2 or less than 1/2, respectively.

Audiometry As the frequencies tested during the audiogram did not match the fundamental frequencies of our stimuli, we used a cubic spline procedure to interpolate or extrapolate hearing thresholds for each stimulus frequency. To account for the log-linear perception of frequency, we fit the cubic spline to hearing thresholds as a function of the log of frequency, rather than frequency in Hz.

Statistical Analysis We analyzed the effect of pitch on sensorimotor timing using a pair of linear mixed models, one to predict percent asynchrony and one to predict relative inter-tap interval. Each model contained fixed slopes for interonset interval, the linear and quadratic effects of pitch height, as well as the interaction between pitch height and IOI. We fit random participant-level intercepts and slopes for interonset interval, under the assumption that participants would differ in their ability to synchronize with and maintain different tempos.

Each model also included hearing threshold as a covariate, as determined from the participant's audiogram. Although we calibrated all tones to be equally loud for participants with normal hearing, individual hearing loss may still cause some pitches to be perceived as quieter than others. As there is some evidence that loudness can influence perceived tempo, with increasing loudness associated with a faster tempo (Boltz, 2011), differences in perceived loudness across our stimuli introduce a potential confound. For example, the prevalence of high frequency hearing loss may cause extremely high pitches to be perceived as quieter, and therefore slower, than other stimuli. By accounting for any hearing loss in our participants, we can separate the effects of pitch and loudness on sensorimotor timing.

To evaluate the significance of all fixed effects, we performed F-tests using the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2017). In cases where we observed a significant effect of pitch height, we determined whether the linear and/or quadratic terms were significant by using lmerTest to perform post hoc t-tests with Holm–Bonferroni correction on the slope estimates.

3.3.2 Results

We begin by discussing the effect of pitch on mean asynchrony during synchronization tapping, and then discuss its effect on the tempo of continuation tapping.

Synchronization Tapping

Figure 3.4A shows mean asynchronies for synchronization tapping to isochronous tones of each combination of pitch and interonset interval, expressed as a percent of that IOI. Larger negative asynchronies indicate earlier tapping relative to the tone being anticipating. Participants tapped earliest when synchronizing to A4 (440 Hz) or A5 (880 Hz) and latest when synchronizing to A2 (110 Hz) or A7 (3520 Hz). Our linear mixed model analysis confirmed that pitch height significantly affected the percent asynchrony of synchronization taps, $F(2, 381.5) = 35.48$, $p < .001$. The effect of pitch was characterized by a significant positive quadratic slope, $t(380.1) = 8.38$, $p_{adj} < .001$, and a nonsignificant linear slope, $t(383.0) = 0.55$, $p_{adj} = .583$, consistent with the mostly symmetrical U-shape of the data in Figure 3.4A. The main effect of IOI was nonsignificant, $F(1, 38.0) = 0.35$, $p = .559$, suggesting that participants tapped at a consistent percent asynchrony regardless of the IOI of the stimulus. The interaction between IOI and pitch was also nonsignificant, $F(2, 380.0) = 1.18$, $p = .310$, indicating that pitch did not differentially affect sensorimotor synchronization at different tempos. Higher (worse) hearing thresholds predicted later tapping on average ($\beta = 0.036$, $SE = 0.028$), but not significantly so, $F(1, 391.3) = 1.61$, $p = .205$.

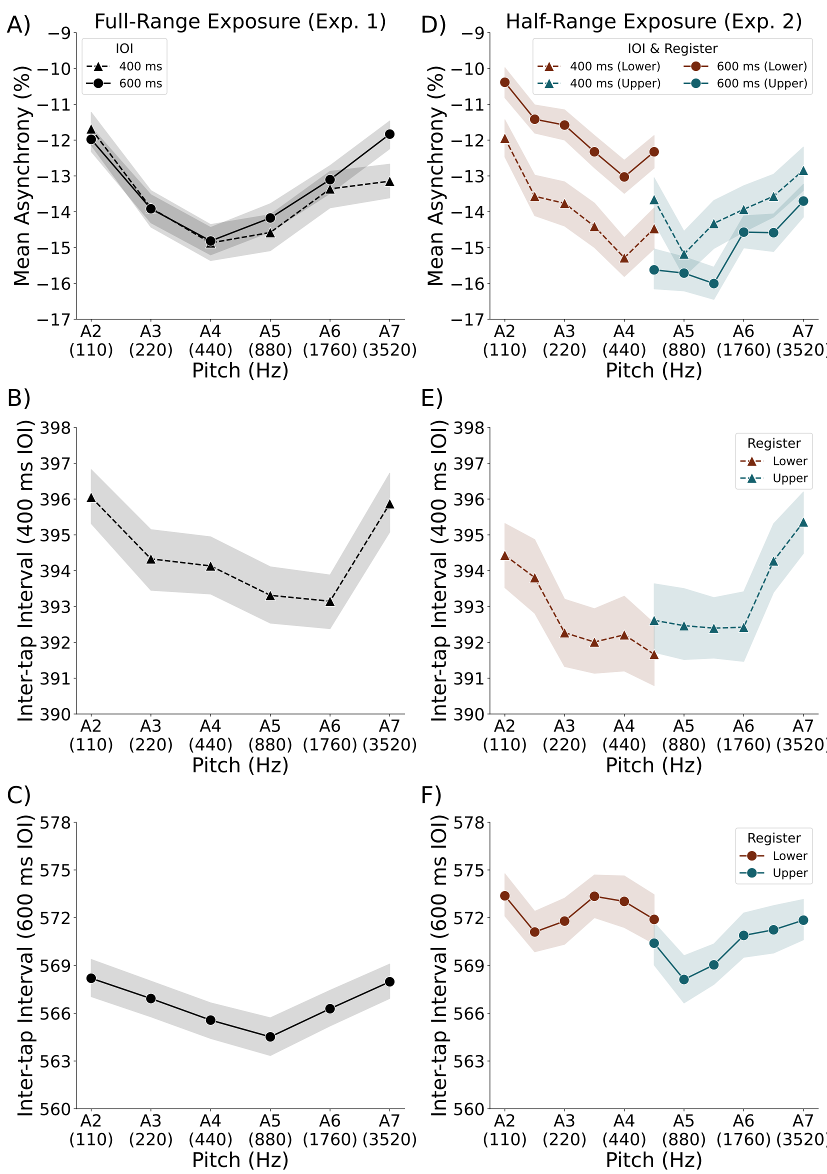

Figure 3.4: Pitch height exerted a U-shaped effect on sensorimotor timing across the frequency range of 110–3520 Hz, regardless of whether participants heard the full five-octave range (Experiment 1; **A–C**) or tones from only one half of the full range (Experiment 2; **D–F**). The latest and slowest timing was consistently produced by both low- and extremely high-pitched tones. **A & D)** Mean asynchrony as a function of the pitch and interonset interval (IOI) of the synchronization tones, expressed as a percent of that IOI. **B & E)** Average number of milliseconds between continuation taps, following synchronization to a 400 ms IOI. **C & F)** Average number of milliseconds between continuation taps, following synchronization to a 600 ms IOI. Shaded bands indicate 95% confidence intervals.

Continuation Tapping

Figures 3.4B–C show average inter-tap intervals during continuation tapping that followed synchronization to tones played at 400 ms or 600 ms intervals, respectively. Note that, although we show raw inter-tap intervals in Figure 3.4 for interpretability, our statistical analysis focused on the ratio between the inter-tap interval and the target IOI. At both tempos, participants tapped slowest after synchronizing to A2 (110 Hz) or A7 (3520 Hz), and fastest after synchronizing to pitches between A4 (440 Hz) and A6 (1760 Hz). Our analysis confirmed that pitch height significantly affected relative inter-tap intervals during continuation tapping, $F(2, 381.3) = 5.79$, $p = .003$. Similar to our asynchrony results, the effect of pitch was characterized by a significant positive quadratic slope, $t(380.1) = 3.40$, $p_{adj} = .002$, and a non-significant linear slope $t(382.6) = 0.07$, $p_{adj} = .946$. Participants tapped slowest after synchronizing to low and extremely high pitches. IOI also significantly affected relative inter-tap interval, $F(1, 38.0) = 45.16$, $p < .001$. Participants tapped faster than the synchronization tones at both tempos, but underestimated the IOI more when asked to match the slower, 600 ms IOI. The interaction between IOI and pitch was not significant, $F(2, 380.0) = 0.10$, $p = .907$. Higher hearing thresholds predicted slower tapping on average ($\beta = 2.4 \times 10^{-4}$, $SE = 1.6 \times 10^{-4}$), but not significantly so, $F(1, 390.1) = 2.22$, $p = .137$.

3.3.3 Discussion

Experiment 1 successfully replicated the findings of Experiments 1–3 from Pazdera and Trainor (2024b) in a synchronization–continuation tapping task. Pitches A2 (110 Hz) and A7 (3520 Hz) produced both the latest synchronization tapping (i.e., smallest negative mean asynchrony; Figure 3.4A) and the slowest continuation tapping (i.e., largest inter-tap intervals; Figure 3.4B–C). Meanwhile, A4 (440 Hz) produced the earliest synchronization tapping, and A5–A6 (880–1760 Hz) produced the fastest continuation tapping. Similarly, Pazdera and Trainor (2024b) observed that participants perceived A2 and A7 as subjectively slowest (Figure 3.1A), while the subjectively fastest pitch varied between A4 and A6 across experiments. Observing the same U-shaped relation between pitch and timing in the absence of an explicit decision-making task suggests that pitch biases timing at the level of perception, rather than decision-making—a distinction that could not be made from our previous study.

3.4 Experiment 2

In Experiment 1, we demonstrated that the U-shaped effect of pitch on perceived tempo observed by Pazdera and Trainor (2024b) also applies to sensorimotor timing. In Experiment 2, we examined whether the linear increase in perceived tempo with increased pitch, previously seen under conditions where only the lower or upper range of pitches were included (Pazdera & Trainor, 2024b, Experiment 5; see Figure 3.1B), would also be observed for sensorimotor responses. Specifically, we presented each participant with pitches from only the lower half or only the upper half of the full 110–3520 Hz range used in Experiment 1.

3.4.1 Methods

Participants

Data for Experiment 2 were collected between May and October 2023, with the COVID-19 safety protocols from Experiment 1 remaining in place for all participants collected before August. Forty-six undergraduate students (34 female, 12 male) from McMaster University completed the experiment for course credit. Ages ranged from 17–24 ($M = 18.3$, $SD = 1.1$). Twenty-two participants (14 female) were randomly assigned to hear lower-register tones only, while the remaining 24 (20 female) were assigned to hear upper-register tones only. An additional six participants completed the experiment, but were excluded (four in the lower-register condition and two in the upper-register condition; see Data Analysis).

Materials

We generated all tones according to the same procedure used in Experiment 1, and all text was displayed in the same font as before.

Apparatus

All hardware was identical to that used in Experiment 1. We updated the Arduino routine to pre-load each new tone immediately after the previous tone stopped playing, instead of immediately before playing the new tone. This change reduced the pre-stimulus "blind spot", during which the Arduino could not read from the tapping pad (see Experiment 1 Data Analysis), from approximately 8 ms to 2 ms. Although the blind spot cannot be eliminated entirely without a multiple-Arduino setup, its impact can be minimized by moving the majority of the audio processing time close to the offbeat, when participants are unlikely to tap. We also added the ability for the Arduino to send the computer information about the peak pressure of each tap and the timestamp of that peak pressure. Although we do not report on the tapping pressure data in the present manuscript, we have included these data in our open-access dataset.

Design

Experiment 2 followed a 2 (Register) × 6 (Pitch Height) × 2 (Interonset Interval) mixed design. Participants were randomly assigned to hear six unique pitches from either the lower half or upper half of the pitch range used in Experiment 1. The pitch of the stimulus varied between trials, with participants assigned to the lower register hearing one of A2 (110 Hz), D♯3 (155.6 Hz), A3 (220 Hz), D♯4 (311.1 Hz), A4 (440 Hz), or D♯5 (622.3 Hz) on each trial. Those assigned to the upper register instead heard one of D♯5 (622.3 Hz), A5 (880 Hz), D♯6 (1244.5 Hz), A6 (1760 Hz), D♯7 (2489.0 Hz), or A7 (3520 Hz) on each trial. The IOI of the synchronization tones again varied between trials, and was either 600 ms (100 BPM) or 400 ms (150 BPM).

Procedure

We again collected an audiogram from each participant via the same procedure as Experiment 1, with the exception that participants assigned to the lower register only received the descending half of the audiogram (from 1000 Hz to 125 Hz) and participants assigned to the upper register only received the ascending half of the audiogram (from 1000 Hz to 8000 Hz). We collected only half of the audiogram in order to limit participants' exposure to stimuli from the register they were not assigned to. The spontaneous motor tempo and synchronization–continuation tasks followed identical procedures to those of Experiment 1, with two exceptions. First, participants heard a different set of six unique pitches across trials depending on which register they were assigned to (see Design). Second, we corrected the positioning of the self-paced breaks, such that there were always 24 trials between breaks.

Data Analysis

We processed our tapping and audiometry data using the same methods as Experiment 1. We excluded five participants for poor synchronization (mean resultant vector length less than .45), and one additional participant for rushing through the continuation tapping task. A further three participants showed anomalous synchronization or continuation tapping behavior for the first one to two blocks of the experiment, but performed the task correctly on all remaining blocks. For these participants, we excluded only the affected blocks. We performed the same statistical analyses as in Experiment 1, with the exception that we fit separate linear mixed models for each of the two register conditions, as each register used a unique set of pitch heights.

3.4.2 Results

We present our synchronization tapping results first for participants assigned to each register, followed by our continuation tapping results for each group.

Synchronization Tapping

Figure 3.4D shows mean asynchronies for each combination of pitch and interonset interval in Experiment 2, when participants were exposed to tones from only one half of the full range of pitches used in Experiment 1. In general, participants assigned to the lower register (110 - 622.3 Hz) tapped earlier while synchronizing to higher pitches, whereas participants assigned to the upper register (622.3 Hz - 3520 Hz) tapped later while synchronizing to higher pitches. Looking across registers in Figure 3.4D, there appears to a reversal in direction just above A4 (440 Hz) or below A5 (880 Hz).

Among participants assigned to the lower register, pitch significantly affected the percent asynchrony of synchronization tapping, $F(2, 220.7) = 20.53$, $p < .001$. The effect of pitch was characterized by both a significant negative linear slope, $t(221.5) = -5.78$, $p_{adj} < .001$, and a significant positive quadratic slope, $t(219.9) = 2.83$, $p_{adj} = .005$, matching the hook-like shape of the lower-register data. Unlike in Experiment 1, tempo significantly affected percent asynchrony in the lower register, $F(1, 22.0) = 5.24$, $p = .032$, such that participants tapped at an earlier phase when synchronizing to a 400 ms IOI, compared to a 600 ms IOI. However, the interaction between tempo and pitch was again nonsignificant, $F(2, 219.9) = 0.50$, $p = .606$, suggesting that pitch affected percent asynchrony similarly across tempos. Higher hearing thresholds predicted later tapping on average ($\beta = 0.029$, $SE = 0.043$), but not significantly so, $F(1, 229.4) = 0.46$, $p = .497$.

Among participants assigned to the upper register, pitch also significantly affected percent asynchrony, $F(2, 240.4) = 12.37$, $p < .001$. The effect of pitch was characterized by both a significant positive linear slope, $t(240.4) = 4.31$, $p_{adj} < .001$, and a significant positive quadratic slope, $t(240.5) = 2.69$, $p_{adj} = .008$, consistent with the check-mark shape of the upper-register data. IOI did not significantly affect percent asynchrony, $F(1, 24.0) = 1.37$, $p = .253$, and the interaction between IOI and pitch was not significant, $F(2, 240.0) = 0.43$, $p = .651$. Higher hearing thresholds significantly predicted later tapping in the upper register ($\beta = 0.093$, $SE = 0.031$), $F(1, 248.2) = 8.95$, $p = .003$.

Continuation Tapping

Figures 3.4E–F show average inter-tap intervals for 400 ms and 600 ms IOI trials, respectively, during Experiment 2. Similar to Experiment 1, participants tapped faster than the target tempo in both the 400 ms and 600 ms IOI conditions, but tapped closer to the correct tempo in the 400 ms IOI condition. Participants assigned to hear lower-register tones generally tapped faster after synchronizing to higher pitched tones, whereas participants assigned to the upper register generally tapped slower after synchronizing to higher pitched tones.

Among participants assigned to the lower register, our linear mixed modeling analysis indicated that pitch significantly affected relative inter-tap intervals during continuation tapping, $F(2, 220.5) = 5.11$, $p = .007$. The effect of pitch was characterized by a significant negative linear slope, $t(221.1) = -3.13$, $p_{adj} = .004$, and a nonsignificant positive quadratic slope $t(220.0) = 0.69$, $p_{adj} = .494$. Within the lower register, participants tapped significantly faster after synchronizing to higher pitches. Tempo also significantly affected relative inter-tap interval, $F(1, 22.0) = 21.00$, $p < .001$, with participants underestimating 600 ms IOIs to a greater extent than 400 ms ones. Although Figures 3.4E–F suggest a clearer effect of pitch at 400 ms than at 600 ms, the interaction between tempo and pitch was not significant, $F(2, 220.0) = 1.94$, $p = .146$. Lastly, higher hearing thresholds predicted significantly faster tapping on average ($\beta = -8.2 \times 10^{-4}$, $SE = 2.3 \times 10^{-4}$), $F(1, 226.8) = 12.49$, $p < .001$.

Among participants assigned to the upper register, pitch also significantly affected relative inter-tap intervals, $F(2, 240.2) = 6.04$, $p = .003$. The effect of pitch was characterized by a significant positive linear slope, $t(240.1) = 3.04$, $p_{adj} = .005$, and a nonsignificant positive quadratic slope $t(240.2) = 1.90$, $p_{adj} = .059$. Opposite to the pattern observed among participants in the lower register, participants in the upper register tapped slower after synchronizing to higher pitches. Tempo again significantly affected relative inter-tap interval, $F(1, 24.0) = 50.51$, $p < .001$, in a pattern consistent with that of participants in the lower register. The interaction between tempo and pitch was nonsignificant, $F(2, 240.0) = 0.39$, $p = .680$. Hearing thresholds did not significantly predict relative inter-tap interval among participants in the upper register ($\beta = 4.1 \times 10^{-5}$, $SE = 1.2 \times 10^{-4}$), $F(1, 242.9) = 0.11$, $p = .739$.

3.4.3 Discussion

Unlike in Experiment 5 of our previous subjective rating study (Pazdera & Trainor, 2024b; Figure 3.1B), exposing participants to tones from only the upper half of the full 110–3520 Hz pitch range did not cause extremely high pitches to be perceived as fast; rather, they were perceived as slower than medium pitches (Figure 3.4D–F). Within the lower register, participants tapped earlier and faster to higher pitches, as expected. Within the upper register, however, participants tapped later and slower to higher pitches, consistent with Experiment 1 (Figure 3.4A–C). Asynchrony also showed significant positive quadratic effects of pitch in both registers, as the relation between pitch and sensorimotor timing reversed direction above 440 Hz in the lower register and below 880 Hz in the upper register. The significant quadratic term further highlights the failure to eliminate the overall U-shaped relation between pitch and perceived timing by restricting the pitch range. The effects of pitch were even similar in scale to those observed in Experiment 1, with mean asynchrony differing across pitches by about 3% of the interonset interval and inter-tap intervals differing by less than 1%.

Whatever factor previously caused the change in subjective tempo ratings for extremely high pitches in exclusively upper-octave contexts (Pazdera & Trainor, 2024b) does not appear to carry over to sensorimotor timing. Accordingly, we conclude that the U-shaped relation between pitch and timing operates at the level of perception and depends on absolute pitch height, rather than relative pitch height within a context.

3.5 General Discussion

Across two synchronization–continuation tapping experiments, we investigated how the pitch of a rhythmic auditory stimulus influences sensorimotor timing. Testing the same five-octave range of pitches as in the perception study by Pazdera and Trainor (2024b), we found the same U-shaped relation between pitch and timing. We observed the earliest and fastest sensorimotor timing for the same range of pitches (between 440–1760 Hz) that a separate group of participants rated as subjectively fastest in our previous study. Below this range, higher pitches were consistently associated with faster perceived timing; above this range, higher pitches were consistently associated with slower perceived timing.

We previously argued based on Pazdera and Trainor (2024b) Experiment 5 that the U-shaped relation depends on pitch height relative to the lowest pitches in a context. We suggested that pitch might be positively correlated with perceived tempo within a range of two to three octaves of baseline—perhaps tied to human vocal ranges (Kuhn et al., 1979)—before reversing direction after exceeding that range.

Experiment 2 of the present experiment instead supports the hypothesis that the U-shaped relation between pitch and tempo perception depends on absolute pitch. Given that we have now observed slow timing at A7 (3520 Hz) in six experiments, the simplest explanation is that the U-shaped effect does depend on absolute pitch, and that the unusual effects in our previous subjective rating study (Figure 3.1B) resulted from another unidentified factor. Further work is needed to determine why we observed a qualitatively different effect in Pazdera and Trainor (2024b) Experiment 5 compared to all six other experiments.

3.5.1 Attenuated Effects on Continuation Tapping

Although pitch significantly affected both mean asynchrony and continuation tapping tempo, we found a larger effect of pitch on the former in both experiments. Mean asynchrony varied across octaves by approximately 3% of the interonset interval of the pacing signal, whereas inter-tap intervals varied by less than 1%. For comparison, our previous subjective rating study found that perceived tempo varied across octaves by about 4–5% (Figure 3.1A; Pazdera & Trainor, 2024b), similar to the scale of the effect on asynchrony in the present study. This pattern of results suggests that one or more factors attenuated the effect of pitch during continuation tapping.

One possible explanation is that providing auditory feedback allowed participants to recognize their own incorrect timing and correct for the majority of the pitch-induced bias. Specifically, when the initial transition from synchronization to continuation happens, the participant can hear the stimulus sequence suddenly change tempo. As the task instructions emphasized keeping the stimuli playing "at the same steady pace," participants should respond by adjusting their taps in such a way as to undo their initial bias.

For example, if a participant perceived a particular 600 ms stimulus as 3% faster than it truly is due to pitch-induced bias, they might begin tapping at 582 ms intervals to match their internal representation of the tempo. However, this will cause the stimulus sequence to speed up from a 600 ms IOI to a 582 ms IOI, which—with a +3% tempo bias—will be perceived as a shift from 582 to 565 ms. If participants are able to recognize this shift as a result of their own timing error, we might expect them to correct the majority of this error over the next few beats (Mates, 1994b, 1994a; Michon, 1967), leaving relatively little effect of pitch on their inter-tap intervals. Accordingly, we may have observed a larger effect during continuation tapping if performed in the absence of auditory feedback. Without feedback, we might expect the participant to simply continue trying to maintain the 582 ms interval they originally perceived, perhaps while also drifting towards their spontaneous motor tempo (Roman et al., 2023; Zamm, Wang, & Palmer, 2018).

Alternatively, our data may indicate that pitch affects the relative phase of neural entrainment to the beat more than its tempo. During an MEG study investigating the effect of pitch change on neural entrainment, Herrmann et al. (2013) found that pitch shifted the relative phase of entrained oscillations, without significantly changing the frequency of those oscillations. Furthermore, a classification analysis they conducted found that pitch-induced phase shifts predicted subjective ratings of tempo change, potentially bridging the gap between neural dynamics and subjective tempo; however, it remains an open question why a shift in neural phase without a change in frequency would be consciously experienced as a faster tempo. Regardless, if we assume that sensorimotor timing reflects underlying neural dynamics, our data might be interpreted as a behavioral consequence of pitch altering the phase of neural entrainment, shifting asynchrony while only slightly altering inter-tap intervals.

Herrmann et al. (2013) proposed that this shift in the relative phase of entrainment is consistent with the brain recruiting oscillators with different spontaneous frequencies to track stimuli with different pitches. A subsequent dynamical systems analysis by Kim and Large (2015) supports the plausibility of this account. Their simulations reveal that when a neural oscillator entrains to a frequency that differs from its own natural frequency, the difference between these two frequencies biases the relative phase of entrainment. Therefore, if the pitch of a rhythmic stimulus affected which neurons in a gradient frequency network (Large et al., 2010, 2015) responded to it, the phase of entrainment might vary with pitch, even though the network remained frequency-locked to the true stimulus tempo. One simulation by Large (2000) tested such a model, in which lower-pitched stimuli were more strongly coupled to neural oscillators with slower natural frequencies. He found that such a model better predicted human synchronization to music than a model in which all pitches were coupled to all oscillators equally. A similar modeling approach may be able to account for the results of the present study, and we believe that future work should focus on incorporating pitch into rhythm perception modeling.

3.5.2 Slower Timing or Better Timing?

One limitation of our study is that our sensorimotor task cannot differentiate slower timing from more accurate timing. This confound arises because humans exhibit a general tendency to tap with a negative mean asynchrony and speed up during continuation tapping (Flach, 2005; Repp, 2005), and these same patterns arise in our data (see Figure 3.4). Therefore, later synchronization tapping also means that participants' taps landed closer to the stimulus onsets. Similarly, slower continuation tapping means that participants deviated less from the target tempo. There has been considerable attention in previous literature regarding a possible superiority of lower pitches for rhythm perception (Hove, Keller, & Krumhansl, 2007; Hove, Marie, Bruce, & Trainor, 2014; Lenc, Keller, Varlet, & Nozaradan, 2018; Repp, 2003) and for inducing spontaneous entrained movement or the urge to move (Cameron et al., 2022; Stupacher, Hove, & Janata, 2016; Varlet, Williams, & Keller, 2020). If this superiority exists, slower timing for A2 (110 Hz) in our study could be due to improved timing accuracy rather than biased timing. [1]

Such an account of our results leaves open the question of why, within the upper register we tested (622.3–3520 Hz), timing was slower/more accurate at higher pitches, rather than lower pitches. If we assume our results were an effect of some pitches producing better timing accuracy than others, it would imply that frequencies as high as 3520 Hz produce synchronization as strong as low frequencies near 110 Hz. At least one simulation study has supported this possibility: Zuk et al. (2018) used a model of the auditory nerve and midbrain to simulate neural synchrony in response to a variety of periodic sounds. In a supplementary analysis, they found a U-shaped relation between stimulus frequency and synchronization strength. Simulated neural synchrony was weakest for tones between 500–1000 Hz, and strongest for tones below 250 Hz or above 2000 Hz (see Zuk et al., 2018, Supplementary Figure 3), which closely aligns with the pattern in our data. We (Pazdera & Trainor, 2024b) previously discounted the strength of neural synchrony as an explanation for the effect

[1] However, see Wojtczak, Mehta, and Oxenham (2017) for a counter-argument that the proposed low-pitch superiority for rhythm might be due to a timing bias, rather than improved accuracy.

of pitch on subjective tempo, due to the elimination of the U-shaped relation when participants only heard tones from the upper register (Figure 3.1B). However, given the persistence of the U-shaped effect on sensorimotor timing under similar conditions in the present study (Figure 3.4D–F), the Zuk et al. (2018) simulation would provide a parsimonious account of our data. One possible explanation for the difference between Experiment 2 of the present study and Experiment 5 of the Pazdera and Trainor (2024b) study is that there are multiple processes by which pitch affects perceived timing. Specifically, a top-down, learned correlation between higher pitches and faster timing may influence decision making, while a bottom-up U-shaped effect influences rhythmic entrainment.

In order to differentiate a learned tempo bias from an intrinsic property of neural dynamics, future research might test whether new correlations between pitch and timing can be learned through novel musical exposure. Boltz (1998, 2011) favored an imputed timing hypothesis, in which the brain learns real-world correlations between pitch and tempo and imposes this expected timing onto the rhythms we hear, similar to a Bayesian prior (e.g., Cannon, 2021; Vuust & Witek, 2014). Previous research suggests that humans implicitly learn both the melodic (Bharucha, 1987; Krumhansl & Kessler, 1982; Trainor & Trehub, 1992, 1994) and rhythmic structure (Jacoby & McDermott, 2017; Jacoby et al., 2024) underlying their native musical culture, and that they can learn a new musical grammar even within a short period of exposure (Loui, Wessel, & Kam, 2010; Loui, 2012; Rohrmeier, Rebuschat, & Cross, 2011). However, the learning of cross-dimensional priors has received little investigation. Therefore, it would be useful to perform a direct test of whether people can learn—and have their timing biased by—novel pitch–timing correlations.

Real-world correlations between pitch and timing in music also require more exploration. One corpus analysis found that ordinally lower parts in polyphonic Western music tend to have fewer notes than the higher parts they accompany, and that higher-pitched instruments tend to play faster during solos than lower-pitched instruments (Broze & Huron, 2013). However, further analysis of cross-cultural corpora would be useful for determining whether the correlations between pitch and timing vary across cultures, and whether there exists either a monotonic or U-shaped relation between pitch and timing that might explain the biases we have observed (Pazdera & Trainor, 2023, 2024b).

3.6 Conclusion

Regardless of whether the effect of pitch on perceived timing derives from a learned perceptual bias or variability in synchronization strength, the present study makes it clear that pitch affects sensorimotor synchronization, and that models of rhythm perception should be extended to account for the influence of pitch.

Chapter 4

Bidirectional interactions of pitch and time

4.1 Preface

Whereas Chapters 2 and 3 focused on mapping out the effect of pitch on perceived tempo, Chapters 4 and 5 examine the inverse effect—whether deviations from isochronous timing can influence perceived pitch. Although a bidirectional effect is central to hypotheses that pitch and timing are integrated in auditory perception, effects of timing on perceived pitch have received considerably less attention in the literature than effects of pitch on time. Chapter 4 attempts to establish the existence of a bidirectional effect between pitch and time using a pair of experiments with similar designs, but inverted independent and dependent variables. Experiment 1 examines the effect of pitch on the perceived mistiming of tones, while Experiment 2 examines whether the mistiming of a tone alters its perceived pitch.

This chapter was published as a conference paper for the *17th International Con-*

ference on Music Perception and Cognition in August 2023, and a publicly available preprint can be accessed at `https://easychair.org/publications/preprint_open/FQ1Q`. In this chapter, we reference the studies reported in Chapter 2 as Pazdera and Trainor (2024b). Due to this chapter's origins as a conference manuscript, the analysis and discussion are presented in an abbreviated format. The subsequent chapter will provide an extended analysis and discussion of the effects of timing on perceived pitch.

4.2 Introduction

The hypothesis that the human brain integrates auditory pitch and timing information into a unified percept is not a new idea. Decades ago, Cohen et al. (1954) proposed that changes in pitch can be understood as a movement through pitch space over time, and that the perception of this movement is biased by lawful associations between time and space. This hypothesis was formalized by Jones (1976), who proposed that the brain represents music as movement through an integrated space of pitch, loudness, and time. In this integrated space, changes along any one dimension lawfully relate to changes in each other dimension. Therefore, information about pitch can inform predictions and expectations about timing, and vice versa. These concepts have continued to gain support in recent years (Boltz, 2017; Henry & McAuley, 2013).

The effects of pitch and pitch change on perceived tempo have been well documented. Typically, higher-pitched speech and music have been associated with faster perceived tempo (e.g., Boltz, 2011; Collier & Hubbard, 1998; Feldstein & Bond, 1981). Ascending pitch has also been associated with perceived acceleration (e.g., Herrmann et al., 2013). However, recent evidence suggests that the relation between pitch and perceived tempo may not always be monotonic. (Pazdera & Trainor, 2024b) observed inverted U-shaped effects of pitch on perceived tempo when participants were exposed to tones across a five-octave range.

In addition to effects on beat-based timing, pitch has also been found to affect single-interval timing and duration judgments. There is some evidence that intervals flanked by one or more high-pitched tones are underestimated in duration (Lake et al., 2014; Pfeuty & Peretz, 2010), whereas the duration of higher-pitched sounds are overestimated (e.g., Cohen et al., 1954).

The inverse question of whether timing can influence the perceived pitch of a sound has received less attention. Early evidence from Madsen and colleagues suggests that tempo changes can drive illusory changes in perceived pitch, such that speeding up is associated with ascending pitch and slowing down is associated with descending pitch (Duke et al., 1988; Geringer & Madsen, 1984; Madsen et al., 1984). However, we are not aware of any research on how changes in single-interval timing influence pitch perception. If our brains integrate pitch and timing information into a unified percept, then we should expect pitch and time to exert a bidirectional influence on one another. We therefore conducted two experiments to investigate the bidirectional nature of pitch-time interactions. Both experiments employed similar two-alternative forced choice tasks. Whereas Experiment 1 tested the biasing effects of pitch height on the perceived mistiming of probes, Experiment 2 tested the biasing effects of probe timing on pitch discrimination.

4.3 Experiment 1

In Experiment 1, we tested the effect of pitch on perceived timing. Participants completed a two-alternative forced choice task in which they listened to a pacing signal consisting of five isochronous beats, continued to track the beat through two silent intervals (see Manning & Schutz, 2013), and then judged whether a final probe played early or late relative to the next beat. In order to separate the effect of the probe's pitch from that of the pacing signal, we assigned participants to one of two task conditions. In the pitched-probe condition, only the pitch of the probe varied across trials. In the pitched-context condition, only the pitch of the pacing signal varied. We also tested participants at two tempos, to assess whether the effect of pitch on perceived timing differs between faster and slower contexts.

4.3.1 Methods

Participants

Fifty-five undergraduate students (4 male, 50 female, 1 nonbinary) from McMaster University completed the experiment for course credit. We randomly assigned 25 participants (22 female) to the pitched-probe condition and 30 (28 female) to the pitched-context condition. Ages ranged from 17–21 years ($M = 18.2$, $SD = 0.7$). An additional thirteen participants completed the experiment but were excluded from analysis due to either failing both attempts at the headphone test ($N = 6$) or completing the task with a negative d' ($N = 7$). We conducted the experiment online between October 2021 and January 2022 due to COVID-19 restrictions.

Materials

Our stimuli consisted of both complex tones and clicks. We generated all complex tones in Python by summing three sinusoidal waves with random phase, including the fundamental frequency (F0) and the first two overtones (F1 and F2), with an amplitude fall-off of 6 dB/octave. The tones were 250 ms in duration and followed a percussive amplitude envelope, consisting of a 10 ms linear rise followed by a 240 ms exponential decay. Clicks were generated using Audacity, and were 50 ms in duration. We then used Audacity to normalize the loudness of all sounds to -14 LUFS to approximate the loudness of other web-based content. To ensure precise interonset interval (IOI) timing, we pre-generated all tone and click sequences as WAV files. We implemented stimulus presentation and response collection in JavaScript using the jsPsych library (de Leeuw, 2015), and hosted our experiment via the web-based platform Pavlovia (`https://pavlovia.org`). We have made all data, code, and materials from both experiments publicly available on the Open Science Framework at `https://osf.io/3ahxe/`.

Procedure

Experiment 1 used a 2 task (pitched probe or pitched context) $\times$ 3 octave (3^{rd}, 5^{th}, or 7^{th}) $\times$ 2 tempo (400 or 600 ms IOI) mixed design. Octave and tempo varied within subjects, whereas the task varied between subjects. On each trial, participants in the pitched-probe condition heard an isochronous series of clicks and judged the timing offset of a subsequent probe tone. Participants in the pitched-context condition instead heard an isochronous, repeating tone and judged the timing offset of a subsequent click. Within each condition, we presented probes at seven unique timing offsets relative to the beat (on-beat, 10% early/late, 20% early/late, or 30% early/late). We repeated each combination of octave, tempo, and offset eight times (once per block), and presented a different pitch class (C, D, D♯, F, F♯, G♯, A, or B) from that octave on each repetition. Practice trials instead used the tones F4, G4, F6, and G6, and always used a 500 ms IOI with a probe offset of 30%.

Participants were instructed to wear headphones during the experiment, and the session began with six trials of a headphone test based on that of Woods et al. (2017). Participants were notified if they failed to answer at least four trials correctly. In this case, they were informed that they may not be able to answer correctly without headphones and were asked to attempt the test again.

Participants next received instructions for the main task. Each trial consisted of a pacing signal followed by a probe. The pacing signal consisted of five isochronous repetitions of a click or tone. Two silent beats followed the pacing signal, and the probe played near the third beat after the signal ended. We instructed participants to keep track of the beat through the silent period in order to determine whether the probe began earlier or later than the next beat should have occurred. Participants were free to choose how to maintain the beat, and provided their trial responses via a key press. There was no time limit to make a response, and the next trial began 1.5 s after the participant responded.

The session consisted of eight practice trials and 336 experimental trials, organized into eight blocks of 42, with self-paced breaks between blocks. Each combination of octave, tempo, and offset appeared once per block. We fully randomized octave and offset within each block, but alternated the tempo every seven trials to limit the difficulty of the task. We provided feedback on practice trials only.

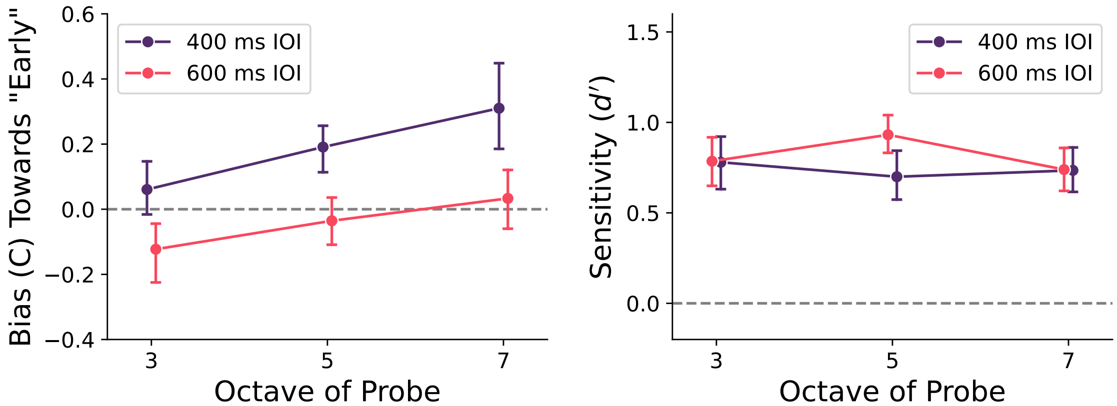

Figure 4.1: Bias (left) and sensitivity (right) of timing offset discrimination in Experiment 1, as a function of tempo and the pitch of the probe tone. Positive values of C indicate a bias towards rating probe tones as earlier. Error bars denote within-subject 95% confidence intervals. Participants rated higher-octave probe tones as earlier than lower-octave probes.

4.3.2 Results

Our primary measures of interest were the bias (C) and sensitivity (d') of participants' offset discrimination judgments. We calculated these measures for each participant at each octave and tempo by considering trials as hits if the participant correctly identified a late probe as late, and false alarms if the participant misidentified an early probe as late. We excluded trials with on-beat probes from analysis, as no correct answer was possible. To prevent hit rates and false alarm rates of 0 and 1, we followed the correction method of Hautus (1995), adding 0.5 to the count of each cell of the contingency table. Under our chosen scoring framework, higher values of C correspond to greater conservatism about rating tones as late (i.e., a bias to rate tones as early).

Pitch of the probe

Figure 4.1 illustrates bias and sensitivity as a function the probe tone's octave and the tempo of the pacing signal in the pitched-probe condition. We analyzed bias via a 3 octave $\times$ 2 tempo repeated measures ANOVA. We observed a large, significant main effect of octave, $F(2, 48) = 6.07$, $p = .004$, $\omega_p^2 = .154$, such that higher-octave probe tones were rated as earlier than lower-octave probes. The main effect of tempo was also significant, with a large effect size, $F(1, 24) = 16.62$, $p < .001$, $\omega_p^2 = .267$. Participants were relatively unbiased in their responses to probe tones that followed a metronome with a 600 ms IOI, but tended to rate probe tones as early when they followed a metronome with a 400 ms IOI. The interaction between octave and tempo was nonsignificant, $F(2, 48) = 0.78$, $p = .466$, $\omega_p^2 = -.003$, suggesting that the probe tone's pitch had a similar effect on its perceived timing regardless of tempo.

We next analyzed sensitivity in the pitched-probe condition via a 3 octave $\times$ 2 tempo repeated measures ANOVA. Both the main effect of octave, $F(2, 48) = 0.52$, $p = .600$, $\omega_p^2 = -.004$, and tempo, $F(2, 24) = 1.79$, $p = .194$, $\omega_p^2 = .007$, were nonsignificant. The interaction between octave and tempo was also nonsignificant, $F(2, 48) = 1.84$, $p = .170$, $\omega_p^2 = .011$. Participants were similarly sensitive to timing offsets regardless of tempo and the octave of the probe tone.

Pitch of the preceding context

Figure 4.2 illustrates bias and sensitivity as a function of the octave and tempo of the tone sequence preceding the probe click in the pitched-context condition. We analyzed bias via a 3 octave $\times$ 2 tempo repeated measures ANOVA. The main effect of octave was small, but significant, $F(2, 58) = 4.07$, $p = .022$, $\omega_p^2 = .036$, and

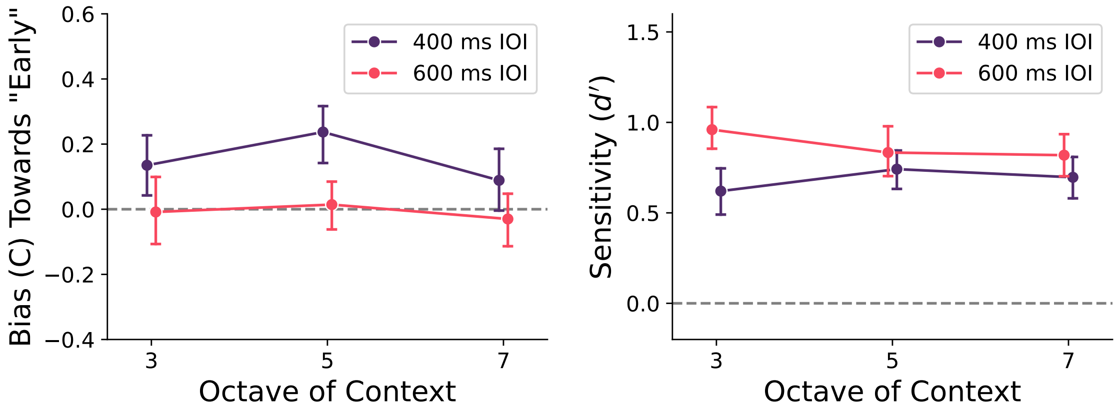

Figure 4.2: Bias (left) and sensitivity (right) of timing offset discrimination in Experiment 1, as a function of tempo and the pitch of the preceding context. Positive values of C indicate a bias to rate probe clicks as earlier. Error bars denote within-subject 95% confidence intervals.

followed an inverted U-shaped pattern with probe clicks perceived as earliest when preceded by a 5^{th} octave sequence. The main effect of tempo was also significant, $F(1, 29) = 4.61$, $p = .040$, $\omega_p^2 = .161$. Similar to our findings in the pitched-probe condition, participants were unbiased in their ratings of probe clicks that followed a sequence of tones with 600 ms IOIs, but tended to perceive clicks following a 400 ms IOI tone sequence as arriving early. The interaction between octave and tempo was again nonsignificant, $F(2, 58) = 1.35$, $p = .268$, $\omega_p^2 = .004$.

Finally, we analyzed sensitivity in the pitched-context condition via a 3 octave × 2 tempo repeated measures ANOVA. The main effect of octave was nonsignificant, $F(2, 58) = 0.12$, $p = .889$, $\omega_p^2 = -.009$, suggesting that participants were similarly sensitive to the timing offset of a probe click regardless of the pitch of the tones preceding it. We did, however, observe a significant main effect of tempo, $F(2, 29) = 7.49$, $p = .010$, $\omega_p^2 = .082$, such that participants were more sensitive to deviations from 600 ms interonset timing than from 400 ms timing. The interaction between octave and tempo was nonsignificant, $F(2, 58) = 3.05$, $p = .055$, $\omega_p^2 = .022$.

4.3.3 Discussion

In Experiment 1, we measured both how a sound's own pitch and the pitch of its context influence its perceived timing. We asked participants to track the beat of an isochronous sequence of clicks or tones, maintain the beat through two silent intervals, and then determine whether a final probe tone or click arrived early or late, relative to that beat. We observed a strong biasing effect of the probe's own pitch on judgments of its timing, as well as a weaker effect of its context. Participants consistently perceived higher-pitched probe tones as earlier than lower-pitched tones (Figure 4.1). This finding is consistent with previous observations associating higher pitch with faster perceived timing in both music (e.g., Boltz, 2011; Collier & Hubbard, 1998) and speech (e.g., Boltz, 2017; Feldstein & Bond, 1981).

In contrast, the effect of the pitch context followed an inverted U-shaped curve, such that probes following middle-octave sequences were perceived as earliest (Figure 4.2). This pattern resembles the inverted U-shaped relation between pitch and perceived tempo identified by Pazdera and Trainor (2024b) whenever participants heard a range of stimuli spanning more than three octaves. However, we anticipated that any bias in the perceived tempo of the pacing signal should exert a bias in the opposite direction on the perceived timing of the probe. For example, if a low-pitched sequence is perceived (and internally represented) as slower than its true tempo, then a subsequent click that occurs on the true beat will arrive earlier than the internal expectation. Thus, if middle-octave sequences are perceived as fastest, we would expect these to be the contexts that make subsequent probes sound the latest – not the earliest. Therefore, we find it unlikely that the effect of the preceding pitch context originated from a biased internal representation of tempo. Rather, our pattern of results is more consistent with pitch biasing the perceived duration of the silent interval between the pacing signal and the probe. This explanation is also consistent with prior observations that perceived interval timing can be biased by the pitch of flanking tones (Lake et al., 2014; Pfeuty & Peretz, 2010).

The biasing effects of pitch were not found to differ across tempos; however, tempo itself did affect timing judgments. Participants in both the pitched-probe and pitched-context versions of the task showed near-zero bias when judging deviations from 600 ms interonset timing. Yet, they showed a general bias to judge probes as early following a pacing signal with 400 ms interonset timing. This pattern is consistent with findings by Vos et al. (1997), who noted a bias for people to perceive the final tones of fast sequences as having sped up.

4.4 Experiment 2

Having observed a strong biasing effect of a tone's own pitch on its perceived timing, our next goal was to determine whether this bias also occurs in reverse. That is, are tones that arrive earlier than expected perceived as higher in pitch than those played later? Experiment 2 addressed this question using a pitch discrimination paradigm, in which participants listened to an isochronous, repeating standard tone and judged whether a subsequent (potentially mistimed) probe tone was shifted higher or lower in pitch. We tested pitch discrimination at two different octaves to assess whether the effect of timing on perceived pitch differs between octaves.

4.4.1 Methods

Participants

Thirty undergraduate students (9 male, 21 female) from McMaster University participated in the study for course credit. Ages ranged from 18–22 years (M = 18.6, SD = 1.1). We conducted the experiment in-lab between March and April 2022 under special COVID-19 safety protocols, as approved by the local research ethics board.

Materials

Our stimuli were complex tones with identical design to Experiment 1, with the exception that we included one additional overtone of the fundamental frequency to improve pitch clarity. We again implemented the experiment using jsPsych (de Leeuw, 2015). The experiment was hosted on Pavlovia, and ran in Google Chrome on a 2011 iMac in our lab. Stimuli were presented at 75 dBA via a pair of Sennheiser HD 201S headphones.

Procedure

The study followed a 3 offset (Early: −15%, On Time: 0%, or Late: +15%) × 2 octave (3rd or 5th) within-subjects design. Within each condition, the probe tone could be presented either higher or lower in pitch relative to the standard. Third-octave standard tones were A3 (220 Hz), with the probe tone shifted by ± 1 Hz. Fifth-octave standard tones were A5 (880 Hz), with the probe tone shifted by ±4 Hz.

Participants completed a pitch discrimination task in which they responded via a key press (up or down arrow) whether a probe tone was higher or lower in pitch than a repeating standard tone. The standard tone repeated six times on each trial at a steady interonset interval of 500 ms, and the probe tone played 425 ms, 500 ms, or 575 ms after the final repetition of the standard. There was no time limit for the participant's response, and the next trial began 1.5 s after the participant responded.

The session consisted of 20 repetitions of each combination of offset, octave, and pitch shift, for a total of 240 trials. We organized trials into four blocks of 60, with self-paced breaks between blocks. Each block consisted of 10 repetitions of each combination of offset and pitch shift, randomly ordered. All trials within a block used standard tones of the same octave to reduce task difficulty, and octave alternated between blocks in an ABAB pattern in which the octave of the first block varied randomly between participants. Four practice trials preceded the main experimental trials and used a standard pitch of A4 (440 Hz) with ±6 Hz shifts and a 0% probe tone offset. Feedback was provided on practice trials only.

4.4.2 Results

We calculated the bias (C) and sensitivity (d') of each participant's pitch discrimination by considering trials as hits when the participant correctly identified a pitch increase, and false alarms when they misidentified a pitch decrease as an increase. We again corrected for extreme hit rates and false alarm rates using the Hautus (1995) method. Under our scoring framework for Experiment 2, higher values of C correspond to greater conservatism about rating tones as high (i.e., a bias to rate tones as low).

Figure 4.3 illustrates bias and sensitivity as a function of probe offset and octave. We analyzed bias via a 3 offset × 2 octave repeated measures ANOVA. Results

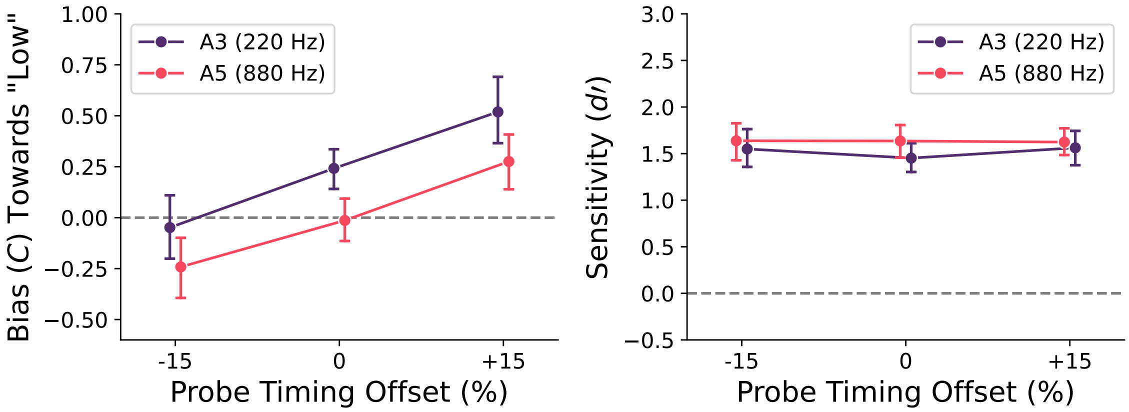

Figure 4.3: Bias (left) and sensitivity (right) of pitch discrimination in Experiment 2. Negative timing offsets indicate early timing and positive offsets indicate late timing. Positive values of C indicate a bias towards rating probe tones as lower in pitch. Error bars denote within-subject 95% confidence intervals. Later timing biased participants to label probe tones as lower in pitch.

indicated a large, significant main effect of probe offset on bias, $F(2, 58) = 14.73$, $p < .001$, $\omega_p^2 = .351$. As the probe timing became later, participants became increasingly biased to label the probe as lower in pitch than its standard. Octave also significantly affected bias, $F(1, 29) = 14.40$, $p = .001$, $\omega_p^2 = .126$. Participants tended to rate probe tones on third-octave trials as lower in pitch than their standards but were relatively unbiased on fifth-octave trials. Offset and octave did not significantly interact, $F(2, 58) = 0.18$, $p = .833$, $\omega_p^2 = -.009$, suggesting that probe timing biased pitch discrimination similarly across octaves.

We next analyzed sensitivity via a 3 offset $\times$ 2 octave repeated measures ANOVA. Neither offset, $F(2, 58) = 0.25$, $p = .784$, $\omega_p^2 = -.009$, nor octave, $F(1, 29) = 0.91$, $p = .347$, $\omega_p^2 = .006$, significantly affected sensitivity, and offset and octave did not significantly interact, $F(2, 58) = 0.23$, $p = .794$, $\omega_p^2 = -.009$.

4.4.3 Discussion

Experiment 1 demonstrated that the pitch of a sound influences its perceived timing. In Experiment 2, we tested whether the timing of a sound also impacts its perceived pitch. We asked participants to listen to an isochronous sequence of standard tones, and to determine whether a final, potentially mistimed tone was higher or lower in pitch. We observed a strong bias to perceive probe tones as lower in pitch, the later they arrived (Figure 4.3). Previous work by Madsen and colleagues has suggested that tempo changes can drive illusory pitch changes in the same direction (Duke et al., 1988; Geringer & Madsen, 1984; Madsen et al., 1984). Our results suggest that even single-interval timing changes can bias perceived pitch. The biasing effects of timing on perceived pitch were similar across both octaves we tested. Octave did, however, directly bias perceived pitch change. At A3 (220 Hz) participants showed a bias to perceive the probe tone as lower in pitch than the standard, whereas at A5 (880 Hz) participants were relatively unbiased at judging pitch change. We are not aware of any previous studies directly comparing biases in pitch discrimination at different octaves.

4.5 General Discussion

Together, our experiments demonstrate strong, bidirectional interactions of pitch and time in auditory perception. Using similar experimental paradigms, we evaluated both the effect of pitch on perceived timing and the effect of timing on perceived pitch. Both experiments revealed a perceptual association between higher pitch and earlier timing. The bidirectional nature of these perceptual biases suggests that pitch and timing are integrated during auditory processing.

The concept that our brains integrate pitch and tempo into a unified percept has been proposed for decades (e.g., Boltz, 2017; Cohen et al., 1954; Henry & McAuley, 2013; Jones, 1976), though models and theories of pitch perception and time perception largely remain independent (however, see Large, 2000, for one example of incorporating pitch into an oscillator model of rhythm perception). We should begin to take seriously the growing body of evidence that our brains integrate pitch and timing information, and work towards integrating pitch and timing models. Doing so will be a critical next step in developing and consolidating our understanding of music perception, and of auditory perception more broadly.

Chapter 5

Timing-induced illusory percepts of pitch

5.1 Preface

In this final data chapter, we begin with an extended analysis of the pitch discrimination experiment presented in Chapter 4. We then present data from a follow-up experiment designed to investigate whether timing-induced biases are stronger under conditions of low pitch discriminability, or in individuals with less precise pitch perception. We conclude with a proposal of how pitch and timing information might be integrated in auditory perception.

Chapter 5 has been made publicly available as a preprint on *PsyArXiv* at `https://osf.io/preprints/psyarxiv/caxsb`. We intend to submit this chapter as a manuscript for an upcoming special collection in *Scientific Reports* entitled "Illusions of the mind." In Chapter 5, we reference the studies reported in Chapter 2 as Pazdera and Trainor (2024b), Chapter 3 as Pazdera and Trainor (2024a), and Chapter 4 as

Pazdera and Trainor (2023).

5.2 Introduction

It has long been suggested that timing and pitch are integrated in auditory perception, such that changes along one dimension influence perceived changes along the other. In particular, Jones (1976) emphasized a principle of proportionality, in which changes in pitch (and loudness) are constrained in magnitude by the time over which they occur; meanwhile, changes in time can only be defined through reference to events that themselves have pitch and loudness. She proposed that the brain integrates the pitch, loudness, and timing of auditory signals into a trajectory through a combined representational space, the structure of which reflects the lawful relations between these three dimensions in the external world. Because of lawful relations like proportionality, movement along any one dimension constrains and biases expectations for movement along each other dimension. For example, early work by Cohen et al. (1954) and more recent work by Henry and McAuley (2009, 2013) has demonstrated a bias to perceive changes in pitch as larger when spaced over a longer interval (often referred to as *tau* effects), and a bias to perceive the timing of notes as if pitch maintained a constant velocity over time (referred to as *kappa* effects).

5.2.1 Integration of Pitch and Timing

Although Jones (1976) focused on the proportionality of the magnitudes of changes in pitch and time, more recent evidence also supports a perceptual link in the *directionality* of changes in pitch and timing. Specifically, there appears to be a perceptual association between pitch increases and temporal acceleration. For example, listeners perceive ascending melodies to be faster (Boltz, 2011) and speeding up more (Collier & Hubbard, 1998) than descending ones, and applying a continuous pitch glide to a melody also induces illusory changes in tempo (Gordon & Ataucusi, 2021). This influence of pitch change on perceived timing extends beyond musical stimuli, as well. Illusions of tempo change have been observed in the perceived modulation rate of frequency-modulated (Herrmann et al., 2013, 2014) and amplitude-modulated (Herrmann & Johnsrude, 2018) tones, with ascending tones perceived as increasing in modulation rate. In speech, it has also been found that listeners are best at recognizing changes in speaking rate and pitch when both features change in the same direction (Boltz, 2017).

Effects of absolute pitch on perceived timing have also been observed. For example, higher-pitched speech (Feldstein & Bond, 1981), melodies (Boltz, 2011), and scales (Collier & Hubbard, 1998) are perceived as faster than lower pitched ones. Additionally, when participants were asked to make early/late judgements about mistimed probes at the end of a rhythmic sequence, Pazdera and Trainor (2023) found that lower-pitched tones were consistently perceived as later than higher ones. Similarly, the P-centers of long, low-pitched tones tend to be later than those of long, high-pitched tones (Danielsen et al., 2019), and single intervals flanked by at least one low-pitched tone have been found to be overestimated (Lake et al., 2014; Pfeuty & Peretz, 2010). New findings from two of our own recent studies suggest that there may be an inverted U-shaped relation between absolute pitch and perceived tempo, in which perceived tempo rises with pitch at lower octaves, but reliably slows above A6 (1760 Hz); however, it remains uncertain whether this U-shaped effect originates from the same mechanism as higher–faster illusions (Pazdera & Trainor, 2024b, 2024a).

There has been considerably less investigation into whether timing also influences perceived pitch; however, if we believe that the brain integrates these two dimensions of sound into a shared representational space, then pitch and timing should bidirectionally influence one another (Boltz, 2017; Jones, 1976). Direct evidence for an effect of timing on perceived pitch is limited, but one collection of studies has observed a biasing effect of tempo changes on perceived pitch changes, such that slowing the tempo of orchestral and band recordings produced perceived decreases in pitch (Duke et al., 1988; Geringer & Madsen, 1984; Madsen et al., 1984). Additional evidence that pitch and timing are at least implicitly associated, if not directly integrated, has been found in musical preference and imagery. When asked to adjust melodies to their preferred tempo, people tend to select faster tempos for higher-pitched music (Tamir-Ostrover & Eitan, 2015), and higher pitch correlates with faster imagined motion in adults (Eitan & Granot, 2006), though not children (Eitan & Tubul, 2010; Kohn & Eitan, 2009, 2016). Auditory Stroop effects have also been found, in which people associate high pitches with the word "fast" and low pitches with the word "slow" (Walker & Smith, 1984). Further study is needed, however, to conclusively support the perceptual integration of pitch and time.

5.2.2 The Present Study

In the present study, we conducted a pair of pitch discrimination experiments to further investigate whether pitch and timing are integrated in auditory perception. Specifically, we tested whether deviations from isochronous timing can induce perceived changes in pitch. In conjunction with previous findings that pitch influences perceived timing, such a reverse-influence of timing on perceived pitch would support the hypothesis that pitch and timing are perceptually integrated and bidirectionally bias one another (Boltz, 2017; Jones, 1976). Although we have previously reported an abbreviated analysis of sensitivity and bias in Experiment 1 of the present study in a conference proceedings (Pazdera & Trainor, 2023), the present manuscript serves to provide a complete analysis and interpretation of this experiment, as well as a follow-up study investigating whether there is a moderating effect of difficulty on timing-induced illusory changes in pitch.

In both experiments, participants listened to an isochronous, repeating standard tone followed by a (potentially) mistimed final tone that was shifted either up or down in pitch. Participants were tasked with determining the direction of the pitch change, and we analyzed whether the timing offset of the probe tone influenced its perceived pitch. We hypothesized that late probe tones would be more likely to be perceived as low-pitched than early ones, given previous evidence for associations between low pitch and slow timing (e.g., Boltz, 2011). We also hypothesized that pitch discrimination would be more sensitive for probe tones played on the beat than for mistimed probes, in line with the principles of Dynamic Attending Theory (Jones & Boltz, 1989; Large & Jones, 1999) and previous empirical evidence (e.g., Chang et al., 2019; Henry & Herrmann, 2014; McAuley & Fromboluti, 2014).

5.3 Experiment 1

In Experiment 1, we tested whether the timing of a probe tone influences its perceived pitch. Participants listened to six isochronous standard tones and rated whether a final probe tone was higher or lower in pitch than the standard. The final tone could arrive early, on the beat, or late, and we evaluated whether these timing deviations biased participants' pitch discrimination responses.

5.3.1 Methods

Participants

We collected data for Experiment 1 between March and April 2022 under special COVID-19 safety protocols, as approved by the local research ethics board, including mask requirements for all participants and experimenters. Thirty undergraduate students (9 male, 21 female) from McMaster University participated in the study for course credit. Ages ranged from 18–22 years, with a mean age of 18.6 ($SD = 1.1$). Of these participants, we excluded five from analysis for failing to perform above chance, as evaluated by a binomial test.

Data Availability

We have made all data, code, and stimuli from both experiments publicly available on the Open Science Framework at `https://osf.io/hrj3t/`, as well as on GitHub at `https://github.com/jpazdera/IllusoryPitch`.

Materials

We used Python to create complex tones with a percussive amplitude envelope by summing four sine waves with random phase, including the fundamental frequency and the first three overtones with an amplitude fall-off of 6 dB/octave. The tones were 250 ms in duration and consisted of a 10 ms linear rise, followed by an exponential decay and 10 ms linear fade. We used Audacity's loudness normalization function, which is based on recommendation ITU-R BS.1770-4 (International Telecommunication Union, 2017), to balance all tones to the same loudness. To ensure precise inter-onset timing, we pre-generated all tone sequences using Python, and played them back as WAV files during the experiment.

Apparatus

Participants completed the study on a 2011 iMac, and we presented stimuli at 75 dBA via a pair of HD 201S Sennheiser headphones. We used the JavaScript library jsPsych (de Leeuw, 2015) to implement stimulus presentation and response collection. Although we conducted the study in person, we used the online platform Pavlovia (`https://pavlovia.org`) to host the experiment, which participants accessed via Google Chrome. The purpose of hosting the experiment on Pavlovia was to enable flexible switching between online and in-person testing in the event that COVID-19 restrictions changed during data collection. Ultimately, however, all participants completed the study in person. We performed all analyses using a combination of Python (version 3.10) and R (version 4.3).

Design

The study followed a 3 probe timing offset (15% Early, On-Beat, 15% Late) $\times$ 2 octave (3^{rd} or 5^{th}) $\times$ 2 pitch shift direction (Up or Down) within-subjects design. Third-octave standard tones were A3 (220 Hz) and fifth-octave standard tones were A5 (880 Hz), with the probe tone shifted by $\pm$7.9 cents, equating to $\pm$1 Hz and $\pm$4 Hz, respectively.

Procedure

Participants completed a pitch discrimination task in which they heard six isochronous repetitions of a standard tone (A3 = 220 Hz or A5 = 880 Hz) followed by a final probe tone. The standard tone always played at an interonset interval of 500 ms, and the probe tone played either 425, 500, or 575 ms after the final repetition of the standard. Following the presentation of the probe, the participant responded via a key press (up or down arrow) whether the probe was higher or lower in pitch than the repeating standard. There was no time limit on their response, and participants were instructed to respond as accurately as possible. The next trial then began 1.5 s post-response.

Trials were administered in four blocks of 60, with each block consisting of 10 repetitions of each of the six combinations of probe timing offset and pitch shift direction, randomly ordered. In order to reduce the difficulty of the task, all trials within a block used standard tones of the same octave, and octave alternated between blocks in an ABAB pattern. The octave of the first block was randomized between participants. Four practice trials preceded the first block, all of which used a standard pitch of A4 (440 Hz), a probe tone that played 500 ms after the final repetition of the standard, and a pitch shift of ±6 Hz (three times larger in cents than the experimental trials). Feedback was provided on the practice trials only. Participants received self-paced breaks between blocks.

Data Analysis

Our primary analysis used a signal detection theory approach to evaluate sensitivity and bias in participants' pitch discrimination. To do so, we marked trials as hits when participants correctly identified a pitch increase, and we marked trials as false alarms when participants misidentified a pitch decrease as an increase. From this scoring, we calculated sensitivity as d' and bias as C (Stanislaw & Todorov, 1999), while correcting for hit rates and false alarm rates of 0 and 1 using the method proposed by Hautus (1995). Specifically, we added 0.5 to the numerator and 1 to the denominator when calculating all hit rates and false alarm rates. Under this labeling scheme, higher values of C indicate greater conservatism about rating the probe tone as higher in pitch than the standard. In other words, positive values of C indicate a bias to rate probes as low-pitched, and negative values of C indicate a bias to rate probes as high-pitched. For each participant, we calculated d' and C separately for each of the six combinations of probe timing offset and octave. We then analyzed these values via a pair of 3 (offset) $\times$ 2 (octave) repeated measures ANOVAs—one for sensitivity and one for bias.

Next, we performed an exploratory analysis to determine whether the biasing effect of probe timing was stronger in individuals with lower sensitivity. Because our planned analysis found an approximately linear effect of timing on bias, we quantified the magnitude of each participant's bias by fitting a linear regression across the C values for their three probe timing conditions, pooling the data from both octaves. The slope of this line indicates the change in bias associated with every 1% delay of the probe tone. A steeper positive slope therefore indicates a stronger overall bias to rate later probe tones as lower in pitch than earlier probe tones, and we refer to this slope as an individual's *timing-induced bias*. We then pooled the data from all conditions to calculate each participant's overall d' sensitivity score, and tested the Pearson correlation between sensitivity and timing-induced bias.

Finally, we tested whether reaction times differed depending on probe timing offset

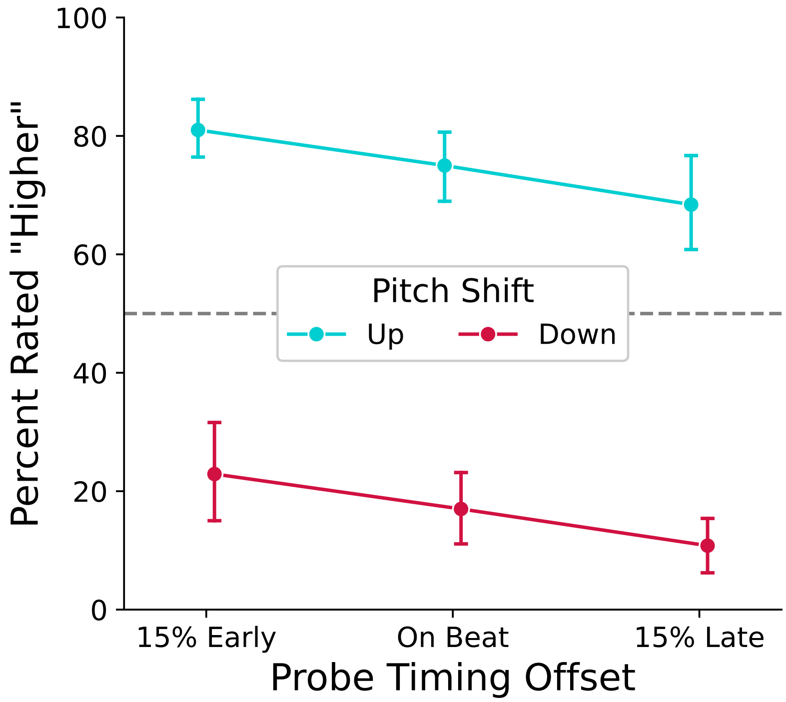

Figure 5.1: Probability of rating probe tones as higher than the repeating standard tone in Experiment 1, depending on the probe tone's timing and true pitch shift direction. Error bars indicate within-subject 95% confidence intervals. In calculating d' and C for all subsequent analyses, we treated correctly-rated pitch increases as hits (upper line) and incorrectly-rated pitch decreases as false alarms (lower line). Hit rates and false alarm rates both decreased as the probe tone became later.

and pitch shift direction. For this analysis, we included only correct responses, while excluding any response with a reaction time slower than 5 seconds (1.6% of correct responses). We then analyzed reaction times via a 3 (offset) × 2 (pitch shift direction) repeated measures ANOVA.

5.3.2 Results

Sensitivity & Bias

Figure 5.1 illustrates the percent of probe tones participants rated as higher in pitch than the standard, as a function of probe timing offset and the true pitch shift direction. These data suggest that the earlier the probe tone played, the more likely participants were to rate it as higher in pitch. In order to separately analyze sensitivity and bias within these ratings, we labeled all trials where participants correctly identified pitch increases (upper line) as hits, and labeled all trials where participants incorrectly responded to pitch decreases (lower line) as false alarms. From these hit rates and false alarm rates, we obtained the sensitivity (d') and bias (C) of participants' pitch discrimination. Figure 5.2 illustrates sensitivity and bias as a function of the octave of the standard tone and the timing offset of the probe tone. Higher values of d' indicate greater discriminability of pitch increases and decreases, while higher values of C indicate a bias towards rating probe tones as lower in pitch than the standard.

We analyzed sensitivity via a 3 (offset) $\times$ 2 (octave) repeated measures ANOVA. Neither probe timing offset, $F(2, 48) = 0.26$, $p = .772$, $\omega_p^2 = -.010$, nor octave, $F(1, 24) = 0.74$, $p = .397$, $\omega_p^2 = .010$, significantly affected sensitivity, and offset and octave did not interact, $F(2, 48) \approx 0.00$, $p > .999$, $\omega_p^2 = -.014$. Participants were similarly sensitive to pitch changes at both octaves, and regardless of whether the probe tone played early, on the beat, or late.

We next analyzed bias via a 3 (offset) $\times$ 2 (octave) repeated measures ANOVA, which indicated significant main effects of both probe timing offset, $F(2, 48) = 10.51$, $p < .001$, $\omega_p^2 = .290$, and octave, $F(1, 24) = 8.99$, $p = .006$, $\omega_p^2 = .133$. The

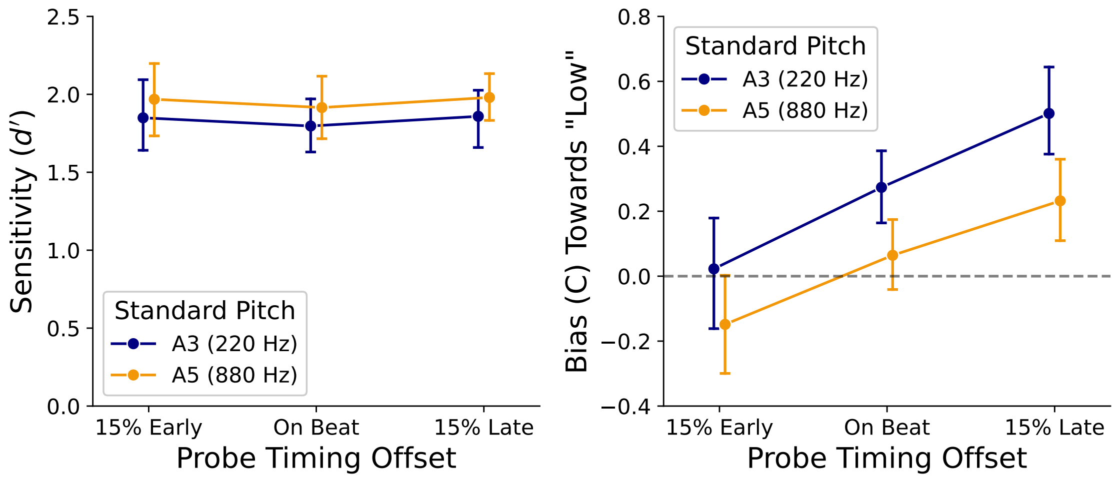

Figure 5.2: Sensitivity (d') and bias (C) of pitch discrimination in Experiment 1, as a function of the octave of the standard tone and the timing offset of the probe tone. Higher values of C indicate a greater bias towards labeling probe tones as lower in pitch than the standard. Error bars denote within-subject 95% confidence intervals. The timing of the probe tone biased participants' pitch perception at both octaves, such that later probes were perceived as lower, without a reduction in discriminability.

interaction between timing offset and octave was not significant, $F(2, 48) = 0.42$, $p = .661$, $\omega_p^2 = -.008$. Post-hoc pairwise t-testing with Holm–Bonferroni correction indicated that the C values for all three probe timing offsets significantly differed from one another, with participants tending to rate later probe tones as lower in pitch, as hypothesized. This pattern was consistent between both octaves we tested; however, participants showed an unexpected main effect of octave such that they were more likely to rate probe tones as lower in pitch when the standard was A3 (220 Hz) than when it was A5 (880 Hz).

Finally, we explored whether the biasing effect of the probe's timing correlated with sensitivity. Figure 5.3 illustrates each participant's overall d' across all trials, paired with the magnitude of their bias to rate later probe tones as lower (formally, the linear slope of their bias across offset conditions, see Data Analysis). Participants

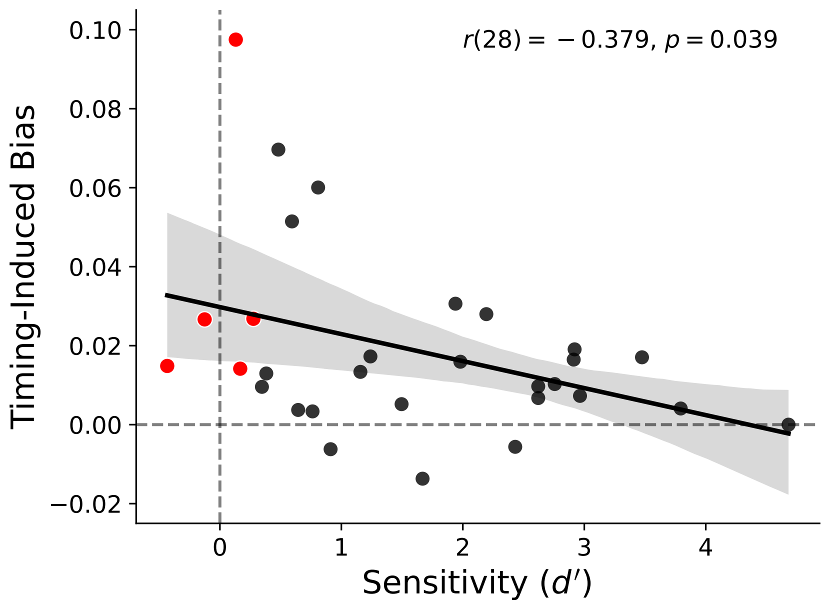

Figure 5.3: Each data point represents one participant's sensitivity (d') and timing-induced bias scores in Experiment 1. Higher timing-induced bias scores indicate a stronger tendency to rate later tones as lower in pitch than the standard. Participants marked in red are those who were excluded from other analyses for failing to perform above-chance. The shaded region indicates the regression line and its 95% confidence interval. Individuals who were less sensitive to pitch changes tended to be more biased by the probe's timing offset.

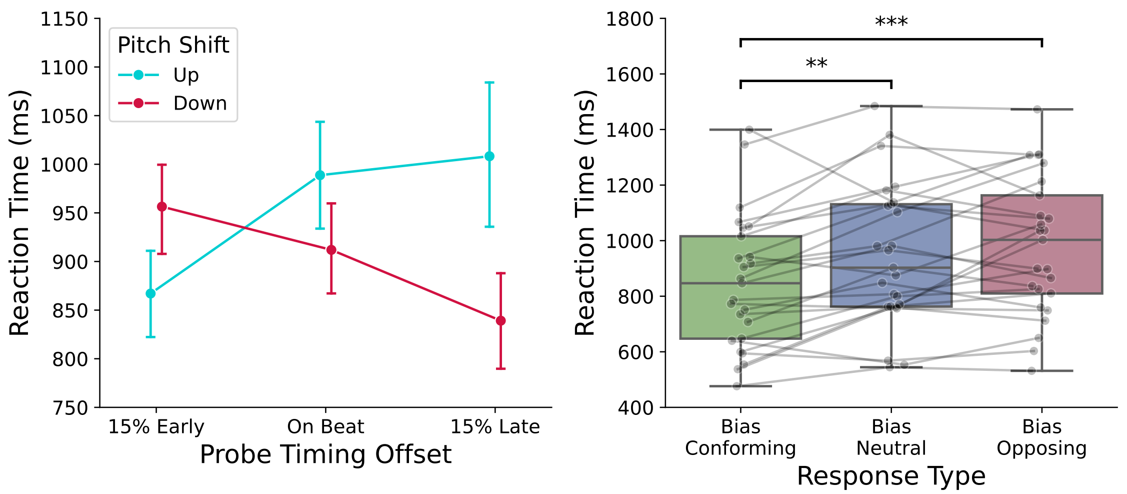

Figure 5.4: Reaction times for correct responses in Experiment 1. **A)** Participants correctly rated pitch increases most quickly when the probe tone was early (blue line), but correctly rated pitch decreases most quickly when the probe was late (pink line). Error bars indicate within-subject 95% confidence intervals (Loftus & Masson, 1994). **B)** Categorizing responses as either bias-conforming (early/high or late/low), bias-neutral (all responses to on-beat probes), or bias-opposing (early/low or late/high) reveals faster reaction times for bias-conforming responses than for both other response categories. Individual data points indicate subject averages for each category.

who failed to perform above chance, and were therefore excluded from our main analyses, are marked in red. Notably, participants with low d' values tended to be highly biased by timing, especially those with $d' < 1$. When including all participants, we observed a moderate negative correlation between sensitivity and timing-induced bias, $r(28) = -.379, p = .039$. Among above-chance performers, this correlation remained moderate in size, but was non-significant, $r(23) = -.330, p = .107$.

Reaction Time

Figure 5.4A illustrates average reaction times for correct pitch discrimination responses, depending on the direction of the pitch shift and the timing offset of the probe tone. A 3 (probe timing offset) × 2 (pitch shift direction) repeated measures ANOVA identified a significant two-way interaction, $F(2, 48) = 13.38$, $p < .001$, $\omega_p^2 = .142$, while neither main effect was significant: probe timing offset, $F(2, 48) = 1.32$, $p = .276$, $\omega_p^2 = .003$, and pitch shift direction, $F(1, 24) = 1.78$, $p = .194$, $\omega_p^2 = .035$. In particular, Figure 5.4A suggests that participants responded correctly most quickly to early, high probes and late, low probes. Therefore, as a post-hoc test of the two-way interaction between timing offset and pitch direction, we recategorized correct responses as either bias-conforming (early/high and late/low probes), bias-opposing (early/low and late/high probes), or bias-neutral (any on-time probe). We next calculated each participant's average reaction time when making each of these three response types, as shown in Figure 5.4B. We then conducted dependent samples t-tests with Holm–Bonferroni correction between each of the three response types. As expected, reaction times for bias-conforming responses ($M = 850$ ms, $SD = 241$ ms) were significantly faster than reaction times for both bias-neutral ($M = 949$ ms, $SD = 255$ ms), $t(24) = -3.88$, $p_{adj} = .001$, and bias-opposing responses ($M = 980$ ms, $SD = 250$ ms), $t(24) = -4.42$, $p_{adj} < .001$. However, bias-opposing responses were not significantly slower than bias-neutral responses, $t(24) = 1.24$, $p_{adj} = .233$.

5.3.3 Discussion

In a pitch discrimination paradigm, we observed a biasing effect of a probe tone's timing on the perception of its pitch. As hypothesized, when a pitch-shifted probe tone played 15% early following six isochronous repetitions of a standard, participants showed a bias to rate the probe as higher in pitch than the standard; meanwhile, when the probe played 15% late, participants showed a bias to rate it as lower in pitch (Figure 5.2). Our results support the idea that pitch and timing are integrated during auditory perception (Jones, 1976). In conjunction with previous findings that higher pitches and ascending pitch sequences are perceived as faster (Boltz, 2011), earlier (Pazdera & Trainor, 2023), or speeding up (Herrmann et al., 2013), our results support a bidirectional influence in which timing can also influence perceived pitch.

To better understand the biasing effect of tone timing on pitch perception, we also analyzed participants' reaction times. We found that correct, bias-conforming responses to early and late probes (i.e., early/high and late/low) were approximately 100 ms faster on average than correct judgments of on-beat probes. In contrast, bias-opposing responses (i.e., early/low and late/high) were not significantly slower than responses to probe tones that played on the beat. We provide a detailed interpretation of this pattern of reaction times in the General Discussion.

The timing-induced bias we identified was not accompanied by a decrease in sensitivity to pitch changes; indeed sensitivity was quite consistent across early, on-beat, and late probe tones, in contrast with our hypothesis that d' would be highest for on-beat probes. One possible explanation for the lack of a dynamic attending-style advantage for on-beat perception (Chang et al., 2019; Henry & Herrmann, 2014; Jones & Boltz, 1989; Large & Jones, 1999; McAuley & Fromboluti, 2014) is that an on-beat sensitivity advantage might require a design in which the majority of probes fall on the beat. In the current design, the probe only played at the "expected" time on one third of trials. Although the average probe timing was on the beat, participants may have learned to spread their attention across the full presentation window due to the high variability in the probe's timing (Large & Jones, 1999). Alternatively, as each trial was only about three seconds in length, trials may have been too short for dynamic attending to emerge. For comparison, previous dynamic attending advantages for pitch perception identified by Chang et al. (2019) were found in sequences lasting 50 seconds.

In addition to time biasing perceived pitch change, participants also unexpectedly showed a bias to rate probe tones as lower than the standard when the standard was A3 (220 Hz), but they were relatively unbiased on average when the standard was A5 (880 Hz; see Figure 5.2). It is possible that our results relate to the pitch class polarization phenomenon identified by Prpic et al. (2016), in which musicians tended to underestimate the pitch class of lower-octave tones and overestimate the pitch of higher-octave tones. However, given substantial differences between our pitch discrimination task and their pitch class identification task, further investigation would be necessary to support a definitive link between our findings.

Lastly, we identified a possible negative correlation between sensitivity and timing-induced bias, such that participants with low sensitivity tended to be more strongly biased by the probe tone's timing (Figure 5.3), but only when we included participants who failed to perform above-chance in the analysis. We designed Experiment 2 to investigate two potential explanations for such a correlation. One possibility is that people may rely on temporal cues as supplemental information when they are uncertain about a pitch change. In this case, we should be able to observe a within-subject effect of task difficulty on timing-induced bias. By varying the size of the pitch shift between trials in Experiment 2, we tested whether individuals would increasingly rely on timing information as pitch changes diminished. Alternatively, individuals with greater pitch sensitivity may simultaneously be better able to differentiate pitch changes from timing changes, allowing them to resist the bias. In Experiment 2, we measured participants' just-noticeable differences for pitch change, and used this measure to calibrate the task difficulty on an individual basis. If greater pitch sensitivity is associated with improved separability of pitch and timing information, then participants with smaller just-noticeable differences should also tend to show weaker timing-induced bias in Experiment 2.

5.4 Experiment 2

In Experiment 2, we followed up on our exploratory finding that individuals with lower sensitivity to pitch change also tended to show stronger timing-related bias. To do so, we created an adaptive-difficulty version of our pitch discrimination task. We first determined each participant's 70.7% just-noticeable pitch difference (JND) via a staircase procedure in which they rated which of two tones was higher in pitch. After obtaining their JND, we presented them with a task similar to Experiment 1, except that the probe tone shifted by a number of cents either equal to their JND (easier condition), or half that number (harder condition). If timing-related bias is stronger in individuals with weaker pitch sensitivity, then we would expect the effect of probe timing offset to positively correlate with JND (as higher JNDs indicate lower sensitivity). Alternatively, or in addition, if timing-related bias increases with task difficulty, then we would expect a stronger effect of probe timing offset in the harder $\frac{1}{2}$JND pitch shift condition than in the easier condition.

5.4.1 Methods

Participants

We collected data for Experiment 2 between February and April 2023, under the same COVID-19 safety protocols as Experiment 1. Twenty-eight undergraduate students (17 female, 11 male) from McMaster University participated for course credit. Ages ranged from 18-22 years ($M = 18.8$, $SD = 1.2$). We excluded one participant from analysis for failing to perform above chance, as determined via a binomial test.

An additional 13 (12 female, 1 male) undergraduate students aged 18-20 years ($M = 18.4$, $SD = 0.6$) completed an alternative version of the task in which all trials were presented at their JND, and these participants were included only in our analysis of whether JND predicts timing-induced bias.

Materials

Tones were constructed via the same procedure as Experiment 1, with the exception that loudness normalization across octaves was not required due to all tones being within 100 cents of A4 (440 Hz).

Apparatus

Participants completed the study on a Windows 10 computer with an Asus Z87-C motherboard, and we presented stimuli at 78 dBA via a set of Escape HP-3868 headphones. We implemented stimulus presentation in Python (version 3.8) using the PsychoPy library (Peirce et al., 2019), and performed all analyses using Python (version 3.10) and R (version 4.3).

Design

The main pitch discrimination task followed a 3 probe timing offset (15% Early, On-Beat, or 15% Late) × 2 difficulty (Easy or Hard) × 2 pitch shift direction (Up or Down) fully within-subjects design. The easier difficulty condition used pitch changes equal to the participant's JND, whereas the harder difficulty condition used pitch changes equal to one half the participant's JND. Pilot testing suggested that with practice, participants became quite good at differentiating pitch changes at their JND, and we found that setting the more difficult condition to be below their initial JND produced desirable levels of performance.

Procedure

The session began with a difficulty calibration task, in which we determined the participant's 70.7% just-noticeable difference for pitch discrimination via an interleaved staircase procedure. On each trial of the calibration task, participants heard a 440 Hz tone followed by a tone slightly higher or lower in pitch than the first, with a 500 ms interonset interval between them. Participants then answered via a key press (1 or 2) whether the first or second tone was higher. A 1.5 s delay followed their response before the next trial began. We used four interleaved staircases in a 2 pitch direction (second tone higher or second tone lower) × 2 initial pitch shift size (1 cent or 25 cents) design. On each trial, we selected one staircase at random to generate the stimuli for that trial. We used a two-down, one-up procedure such that two consecutive correct answers on trials generated by the same staircase increased the difficulty of the next trial generated by that staircase, reducing the number of cents by which the tones differed (to a minimum of 0); meanwhile, a single incorrect answer reduced the difficulty of the next trial generated by that staircase, increasing the number of cents by which the tones differed (to a maximum of 100). Initially, difficulty changed by 8 cents at a time, and this step size halved after every two reversals in difficulty on a per-staircase basis, to a minimum step size of 1 cent. Each staircase ended after eight reversals in difficulty. After all four staircases had ended, we calculated the participant's JND as the average pitch shift size of the last four reversals from each staircase.

We next used the JND obtained from the calibration task to generate probe tones that were a number of cents above and below the standard tone (A4) equal to that threshold, as well as probe tones that were above and below the standard tone by one half the JND. Participants then completed a pitch discrimination task that followed the same procedure as Experiment 1, with the exception that the standard tones were always A4 and the size of the pitch difference between the standard and probe (JND or $\frac{1}{2}$JND) varied across trials. Trials were again organized into four blocks of 60 separated by breaks, with each combination of probe timing offset, difficulty (pitch shift size), and pitch shift direction presented five times per block in a fully randomized order. Four practice trials with a pitch shift size of four times the JND preceded the first block. Feedback was given on the practice trials only.

Data Analysis

We calculated participants' sensitivity and bias in each condition in the form of d' and C, respectively, using the same methods as Experiment 1 (Hautus, 1995; Stanislaw & Todorov, 1999). To confirm that our difficulty manipulation affected sensitivity as intended, we first analyzed d' via a 2 (difficulty) $\times$ 3 (probe timing offset) repeated measures ANOVA. Next, to assess whether difficulty affected the strength of the later–lower timing bias on pitch perception, we quantified timing-induced bias in a similar manner to Experiment 1. Specifically, for each participant and each difficulty, we fit linear models across the C values for the three probe timing offset conditions. As before, the slope of this line quantifies the expected change in bias with each 1% delay in the timing of the probe tone, which we refer to as the *timing-induced bias*. We then compared the timing-induced bias values from the two shift size conditions using a paired-samples t-test. To determine whether individuals with more sensitive pitch perception were less biased by timing, we calculated the Pearson correlation between participants' JNDs and their timing-induced bias, specifically for trials presented at their JND (the easy condition).

5.4.2 Results

Sensitivity & Bias

We first assessed whether our difficulty manipulation produced lower levels of sensitivity on trials where the pitch shift size was $\frac{1}{2}$JND than on trials where it was equal to their JND. Figure 5.5 illustrates d' for each combination of difficulty and probe timing offset. A 2 (difficulty) x 3 (probe timing offset) repeated measures ANOVA identified a large, significant main effect of difficulty on d', $F(1, 26) = 49.97$, $p < .001$, $\omega_p^2 = .469$, a non-significant main effect of of probe timing offset, $F(2, 52) = 1.59$, $p = .214$, $\omega_p^2 = .005$, and a non-significant interaction, $F(2, 52) = 0.18$, $p = .836$, $\omega_p^2 = -.010$. Smaller pitch shifts were significantly less discriminable than larger pitch shifts, confirming that our difficulty manipulation was successful. Furthermore, consistent with Experiment 1, participants were similarly sensitive to pitch changes

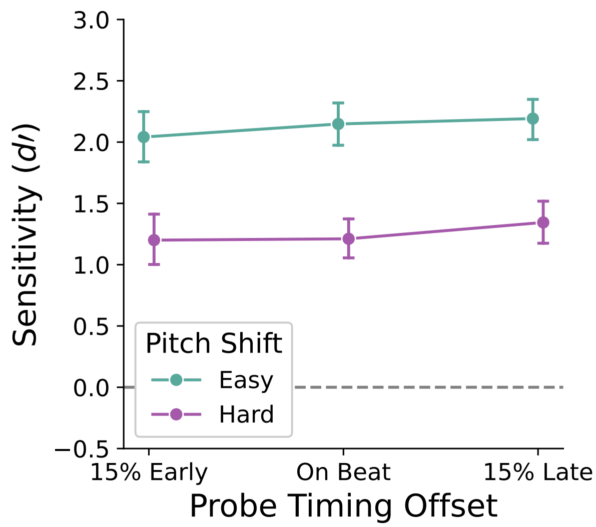

Figure 5.5: Pitch discrimination performance in Experiment 2, based on the size of the pitch shift and the timing of the probe tone. Error bars indicate within-subject 95% confidence intervals. Increasing the difficulty of the task by reducing the size of the pitch shift successfully reduced d' evenly across probe timing conditions.

regardless of whether the probe played early, late, or on the beat.

Having confirmed that our difficulty manipulation impacted pitch discriminability, we next tested whether difficulty affected participants' tendency to rate later probes as lower. Figure 5.6A illustrates C as a function of difficulty and the probe's timing offset. A 2 (difficulty) x 3 (probe timing offset) repeated measures ANOVA identified significant main effects of difficulty, $F(1, 26) = 9.88$, $p = .004$, $\omega_p^2 = .152$, and probe timing offset, $F(2, 52) = 4.67$, $p = .014$, $\omega_p^2 = .067$, as well as a significant two-way interaction, $F(2, 52) = 3.72$, $p = .031$, $\omega_p^2 = .032$. The main effect of difficulty was such that participants showed an overall bias to rate larger pitch shifts as a decrease and smaller pitch shifts as an increase. With respect to the effect of the probe's timing, post-hoc pairwise t-tests with Holm–Bonferroni correction found that late probes were rated as significantly lower than early and on-beat probes. To determine

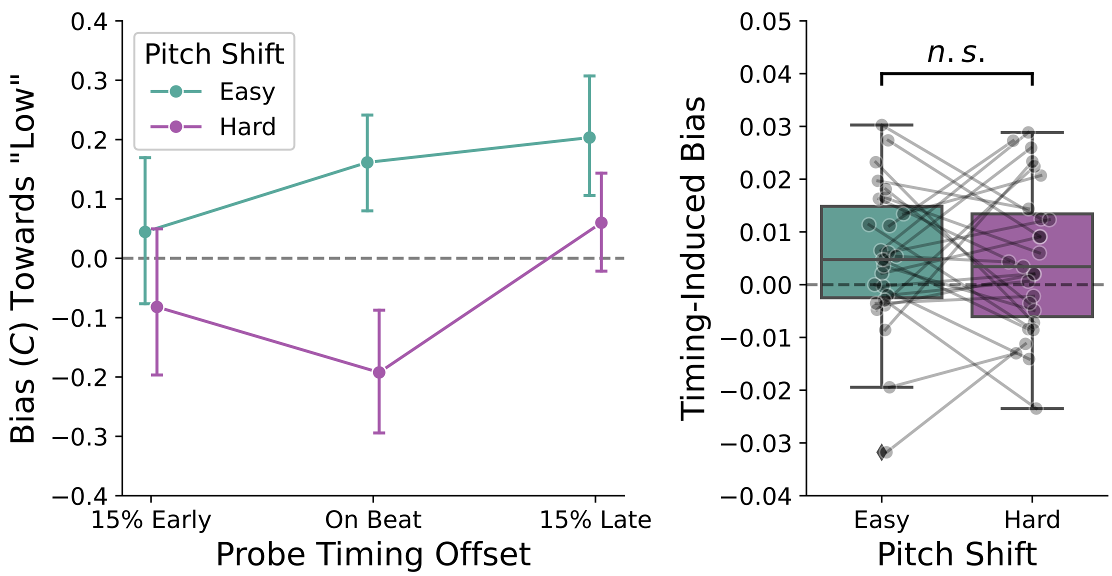

Figure 5.6: Bias in pitch discrimination as a function of the probe tone's timing and the size of the pitch shift in Experiment 2. **A)** Average bias (C) towards rating probe tones as lower than the standard in each condition. Error bars indicate within-subject 95% confidence intervals. Late tones were more likely to be rated as low-pitched than early and on-beat tones. **B)** Data points indicate the linear effect of probe timing offset on bias for each participant and each difficulty condition. Reducing the size of the pitch shift did not strengthen timing-induced bias.

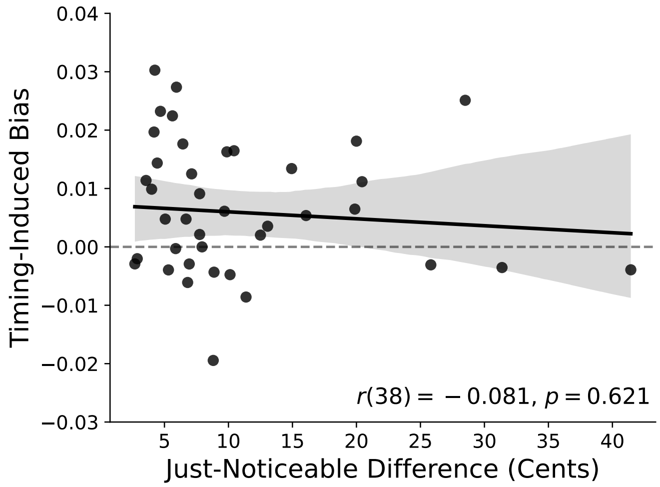

Figure 5.7: Data points indicate each participant's just-noticeable pitch difference in cents, paired with their timing induced bias in Experiment 2. Higher just-noticeable differences indicate less sensitive pitch perception, while greater timing-induced bias indicates an stronger tendency to rate later tones as lower. The shaded region indicates the regression line and its 95% confidence interval. Participants in Experiment 2 were similarly biased by the probe tone's timing regardless of their pitch sensitivity.

whether the two-way interaction matched our hypothesis that timing-induced bias would be stronger when the pitch shift was smaller, we compared timing-induced bias between pitch shift sizes using a dependent samples t-test. According to our hypothesis, timing induced bias should be more positive in the JND condition than the $\frac{1}{2}$JND condition; however, this was not the case, $t(26) = 0.18$, $p = .860$. Rather, the two-way interaction can be accounted for by the difference in C being significantly larger between difficulty conditions when the probe played on the beat than when it played late, $t(26) = 2.57$, $p = .016$.

Just-Noticeable Differences & Timing-Induced Bias

Figure 5.7 illustrates each participant's 70.7% just-noticeable pitch difference alongside the timing-induced bias they exhibited on pitch discrimination trials presented at their JND. A positive correlation would indicate that the biasing effects of probe timing were stronger among participants with less sensitive pitch perception (consistent with Figure 5.3 from Experiment 1), after accounting for task difficulty. Instead, we observed a weak and non-significant negative correlation, $r(38) = -.081$, $p = .621$, suggesting that timing biased participants' pitch perception similarly regardless of their sensitivity to pitch differences.

5.4.3 Discussion

In Experiment 2 we investigated whether the biasing effects of timing on pitch perception vary in strength according to task difficulty and/or individual pitch sensitivity. We conducted an adaptive-difficulty pitch discrimination task calibrated to each person's just noticeable pitch difference. Although we replicated the bias to perceive later probe tones as lower in pitch, we did not find evidence that this bias strengthens when pitch changes are made less discriminable by reducing the size of the change. Timing-induced bias was similarly strong when the pitch change was equal to the participant's JND as when it was half that size (Figure 5.6). We also did not find evidence that the strength of the bias correlated with JND (Figure 5.7). Participants were similarly influenced by the timing of the probe regardless of the precision of their pitch perception. Therefore, neither of these factors appear to account for the sensitivity–bias correlation in Experiment 1.

5.5 General Discussion

Across two pitch discrimination experiments we observed a biasing effect of early versus late tone timing on perceived pitch. Later timing resulted in lower perceived pitch without an impact on discriminability (Figures 5.1–5.2 and 5.5–5.6). The strength of this illusion was not found to depend on task difficulty (the size of the pitch difference between standard and probe tones; Figure 5.6), nor individual differences in sensitivity to pitch changes (measured as just-noticeable difference; Figure 5.7). Alongside previous findings that ascending pitch produces illusions of speeding up (e.g., Boltz, 2011; Collier & Hubbard, 1998; Herrmann et al., 2013), our results provide evidence that pitch and timing bidirectionally influence one another in auditory perception. This bidirectional influence is consistent with accounts that suggest pitch and timing are perceptually integrated (Boltz, 2017; Jones, 1976).

This type of cue integration has often been framed as a Bayesian inference problem (Kersten, Mamassian, & Yuille, 2004; Knill & Richards, 1996; Knill & Pouget, 2004; Vilares & Kording, 2011; Vincent, 2015), and we can apply a similar explanation here. The fundamental idea behind Bayesian models of perception is that the brain needs to infer the state of the surrounding environment based only on noisy sensory information, in conjunction with learned priors regarding the statistical structure of the world. Due to the stochastic nature of neural activity, there are many different states of the world that can produce any given pattern of sensory activation; therefore, incorporating prior knowledge about world structure helps narrow down which of these possible world states generated any given pattern of sensory activity. Optimal Bayesian inference has previously been used to explain illusions in perceived visual (Weiss, Simoncelli, & Adelson, 2002) and tactile (Goldreich & Tong, 2013) motion, and Bayesian/predictive coding accounts of music perception have also emerged within the last decade (Cannon, 2021; Koelsch et al., 2019; Vuust & Witek, 2014).

5.5.1 Auditory Cue Integration as Bayesian Inference

What Jones (1976) described as a "lawful natural relationship" between changes in pitch and time is precisely the type of world structure that might be incorporated into a perceptual prior. Her hypothesis of expected proportionality between changes in pitch and time can be understood as a prior on the velocity of pitch motion (see also Henry & McAuley, 2009, 2013). In nature, larger changes in pitch are statistically more likely to take place over a longer period of time, and so movement along one auditory dimension (pitch or time) does carry information about movement in the other. Bayesian statistics provides a formal mathematical description of *how much* information each dimension provides about the other (Friedman et al., 2013; Genewein, Hez, Razzaghpanah, & Braun, 2015), but the fundamental ideas align closely with Jones (1976). Lawful relations in nature between changes in pitch and time make it possible to partially infer the magnitudes of pitch changes based on elapsed time, and to partially infer the interval between two sounds from the magnitude of pitch change between them. Therefore, when the brain reconstructs the state of the external world based on a combination of incoming noisy sensory data and prior expectations, it is statistically optimal to bias perception towards those priors.

In the context of the present study, we observed a biasing effect of the *direction* of temporal change (rather than magnitude) on the direction of perceived pitch change. This effect might similarly be explained by a prior expectation for a positive correlation between directional changes in pitch and timing. In our task, the probe tone's change in timing is much larger than the change in pitch. The timing change is therefore much easier to detect and, in cases when the probe arrives late, can even be known before the tone plays. Therefore, we can assume the brain is able to infer the directional change in timing faster than the directional change in pitch. This inference allows prior knowledge of correlations between timing changes and pitch changes to inform the ongoing inference about the pitch change. For example, if speeding up is correlated with increasing pitch in nature, then a tone that plays early is more likely *a priori* to have increased in pitch, rather than decreased. Boltz (2011, 2017) has previously argued that there is reason to expect such a real-world correlation to exist, as both higher pitch and faster tempo are higher-energy states. For example objects tend to generate higher-pitched sounds as they speed up, and pitch and tempo might similarly be expected to covary in speech alongside changes in arousal (e.g., Black, 1961). Broze and Huron (2013) have also noted that lower instruments tend to be larger and have slower attack times, constraining how quickly they can be played relative to similar instruments of higher pitch and smaller size. However, a full cross-cultural investigation of pitch–tempo correlations in speech and music will be necessary to conclusively demonstrate a lawful relationship in directionality.

5.5.2 Faster Bias-Conforming Responses

We have established that a Bayesian prior for pitch increasing during acceleration and decreasing during deceleration might explain the biases observed in the present study. But what of our reaction time results? We found that bias-conforming responses to early and late probes were about 100 ms faster than responses to on-beat probes. Bias-opposing responses were slightly slower on average than responses to on-beat probes, but not significantly so (Figure 5.4). We believe this pattern of results can also be explained within the Bayesian framework discussed above, if reliance on the prior allows for faster inference of off-beat pitch changes by acting as an extra cue to the pitch change. All responses to on-beat probes and all correct bias-opposing responses must have been generated based on a slower analysis of the pitch-related sensory activation (or by random guessing). For on-beat probes, this is the case because there is no change in timing to use as a supplemental cue (although a small change in timing could still be incorrectly inferred). For correct bias-opposing responses this is the case because reliance on the prior only biases inference towards an incorrect response. In contrast, correct bias-conforming responses may have either been based on this slower analysis of the pitch-related sensory activity, or through a faster inference process informed by the tone's timing. In this way, correct bias-opposing and bias-neutral responses may be generated by similar processes, while bias-conforming responses sometimes follow a faster, more biased process (for additional discussion of how Bayesian inference may translate to a decision process, see Dunovan, Tremel, & Wheeler, 2014).

Alternatively, fast responses may derive from heightened attention rather than heuristic processing. As we did not restrict response times in our study, we cannot distinguish whether faster responses originated from heuristic processing or differences in attention. In future work, it may be insightful to vary the amount of time participants are given to respond, as one could assess whether timing-induced bias is stronger when participants are forced to rely more heavily on prior expectations to make speeded responses. Attention might also be manipulated through the use of a distractor task (e.g., Herrmann & Johnsrude, 2018).

5.5.3 Difficulty and Perceptual Sensitivity

A Bayesian account might also explain why our selected difficulty manipulation in Experiment 2 (varying the size of the pitch change) did not change the extent to which timing influenced perceived pitch. In multi-cue Bayesian inference, the relative weighting of the cues is proportional to the precision of the estimates that can be made from them (Friedman et al., 2013). The pitch of the probe tone was equally clear in our easy and difficult conditions—the pitch change was just smaller, making it harder to make a discrimination response given a fixed level of sensory precision. From a signal detection theory framework, discriminability in the form of d' is the number of standard deviations between the means of the distributions for the two categories of stimulus. Therefore, one could manipulate discriminability either by moving the two distributions closer together as we did in Experiment 2, or by degrading the stimuli to increase the standard deviation of the distributions. Future research should test whether timing-induced bias increases when pitch changes are degraded rather than made smaller, for example through spectral smearing (Baer & Moore, 1993), as this should theoretically reduce the weighting of spectral information relative to temporal in a Bayesian integration process.

For the same reason, we might expect individuals with less precise pitch perception to down-weight pitch-related sensory information and up-weight temporal information. In Experiment 1, we did observe a negative correlation between d' and timing-induced bias that conforms to this prediction (Figure 5.3); however, we found a near-zero correlation between just-noticeable difference and timing-induced bias during Experiment 2, when difficulty was calibrated at an individual level (Figure 5.7). One reason we might not see a negative correlation between pitch sensitivity and timing-induced bias is if pitch sensitivity correlates positively with temporal sensitivity, such that individuals with poor pitch sensitivity also have poor temporal sensitivity (e.g., Sares, Foster, Allen, & Hyde, 2018). Alternatively, it is possible that timing-induced bias may only significantly increase when pitch sensitivity becomes too poor to perform above chance. As bias tended to be high in participants who failed to perform above chance in Experiment 1 (Figure 5.3), it is possible that the correlation was driven by some participants explicitly using timing as a cue due to the task being too difficult for them to detect any pitch changes at all. In situations where participants can only detect one feature of a stimulus changing, they may conceivably resort to making judgments based on that feature, even if it is not the feature to which they were instructed to attend (Feldstein & Bond, 1981). The individualized difficulty calibration in Experiment 2 may have alleviated this issue and eliminated any correlation between sensitivity and timing-induced bias.

5.5.4 Rhythmic Deviation or Foreperiod Effect?

Although we used deviation from a rhythmic context to manipulate perceived pitch in the present study, it is possible that timing-induced bias depends on the time elapsed since the end (or beginning) of the previous note, rather than on a note's phase within the rhythmic context. This alternative explanation could be tested using a pitch discrimination task in which a single standard tone plays on each trial, followed by a variable delay (i.e., foreperiod) before the onset of the probe tone. Two recent studies by Herbst and Obleser (2017, 2019) implemented a design similar to this without testing the effects of foreperiod duration on bias, as reaction time and accuracy have typically been the focus of foreperiod analyses. We believe our present results support the addition of tests for foreperiod effects on bias in future studies of pitch discrimination. If shorter foreperiods in the absence of a rhythmic context result in higher perceived pitch, our present results might be better explained not by a learned prior on pitch-timing correlations, but rather by an effect in which residual neural activity from one tone exerts a decaying pitch bias on the perception of the next. However, given our previous findings that pitch influences perceived mistiming in the same direction that mistiming influences perceived pitch (Pazdera & Trainor, 2023), we believe that a learned correlation is more likely.

5.6 Conclusion

The present study demonstrates that a tone's timing within a rhythmic context can alter the perception of its pitch. These timing-induced illusory pitch changes—alongside previous evidence for pitch-induced illusory tempo changes—support the long-standing hypothesis that the brain integrates pitch and timing during auditory perception.

Chapter 6

General Discussion

The goal of this thesis was to answer several empirical questions that need addressing before pitch can be incorporated into models of rhythm perception. Although substantial previous evidence has suggested that ascending pitch and higher absolute pitch are associated with faster perceived tempo (see Table 6.1), it has been unknown: 1) whether this higher–faster association generalizes across more than two octaves and above 1000 Hz, 2) whether pitch influences motor tempo in the same pattern that it affects subjective tempo ratings, and 3) whether changes in timing can also create illusory changes in pitch. Across Chapters 2–5, I presented data from ten new experiments that answer these questions. In this final chapter, I review my most important findings and remaining open questions, discuss the theoretical implications of my results and how they may inform future modeling work, and suggest practical applications for the present body of research.

IV	DV	Relation	Relevant Studies	Chapters
Pitch	Tempo	↗ ↗ ↗ ↗ ∩, ↗ ∩	Boltz (2011, 2017) Collier & Hubbard (1998) Feldstein & Bond (1981) Pfeuty & Peretz (2010) **Pazdera & Trainor (2024a)** **Pazdera & Trainor (2024b)**	 (Ch. 2) (Ch. 3)
Pitch	ΔTempo	↗ ↗	Collier & Hubbard (1998) **Pazdera & Trainor (2023)**	 (Ch. 4.3)
ΔPitch	Tempo	↗ ↗	Boltz (2011) Collier & Hubbard (1998)	
ΔPitch	ΔTempo	↗ ↗ ↗ ↗ ↗ ↗	Boasson & Granot (2012) Collier & Hubbard (1998) Gordon & Ataucusi (2021) Herrmann & Johnsrude (2018) Herrmann et al. (2013, 2014) Lake et al. (2014)	
Tempo	Pitch	↗ ↗ ↗ ↗	Boltz (2017) Duke et al. (1988) Geringer & Madsen (1984) Madsen et al. (1984)	
Tempo	ΔPitch	?		
ΔTempo	Pitch	?		
ΔTempo	ΔPitch	↗ ↗	**Pazdera & Trainor (2023)** **Pazdera, Rinaldi, & Trainor (2024)**	(Ch. 4.4) (Ch. 5)

Table 6.1: Revised summary of evidence for interactions between pitch and tempo in auditory perception. (Δ = change in feature, ↗= higher–faster relation, ∩ = negative quadratic relation, ? = untested)

6.1 Summary of Findings

In Chapter 2, I tested the generalizability of higher–faster relations across a five-octave range, extending to pitches nearly two octaves higher than has previously been tested. In doing so, I discovered that the relation between pitch height and perceived tempo is not monotonic under certain circumstances, contradicting previous assumptions that higher sounds are always perceived as faster. Instead, I found an inverted U-shaped (negative quadratic) relation in which perceived tempo increased between A2 (110 Hz) and A4 (440 Hz) reached a maximum between A4 and A6 (1760 Hz), and decreased with pitch beyond that limit. I observed this nonmonotonicity regardless of tone duration (50 ms or 200 ms), timbre (piano tone or synthetic complex tone), and tempo (between 60 and 200 BPM), and synchronous tapping only slightly attenuated the effect. However, the perception of extremely high pitches as slow was replaced by the classic higher–faster bias when participants were exposed only to pitches D♯5 (622.3 Hz) and above. To my knowledge, these data constitute the first evidence for a nonmonotonic relation between pitch and perceived tempo.

In Chapter 3, I used a pair of sensorimotor synchronization experiments to test whether this quadratic relation between pitch and perceived tempo extends to motor timing, and whether the association between extremely high pitches and slow timing can be inverted by exposing people to high-pitched tones only. Consistent with the subjective rating data from Chapter 2, participants tapped earliest and fastest when synchronization–continuation tapping to pitches between A4 and A6. However, participants tapped relatively late and slow for pitches above A6, regardless of whether they only heard D♯5 and above or the full five-octave range. Based on these results, I argued that the quadratic relation between pitch and perceived tempo most likely arises at the level of perception, upstream of the motor system. Furthermore, I proposed that this effect may have separate origins from the tendency to subjectively rate higher pitches as faster, given that the mediating effect of context observed Chapter 2 did not extend to motor tempo.

In Chapter 4, I introduced the idea that effects originating from the integration of pitch and timing cues should be bidirectional. That is, if pitch influences perceived timing, timing should influence perceived pitch in the same direction. By pairing a temporal discrimination experiment with a pitch discrimination experiment of similar design, I provided evidence for bidirectional interactions of pitch and time. Specifically, I showed that mistimed tones were perceived as earlier, the higher they were in pitch, while also being perceived as higher in pitch, the earlier their timing. The effects of pitch on perceived mistiming did not demonstrate the inverted U-shaped relation I observed in tempo ratings and motor timing, despite the temporal discrimination task using a range of pitches (C3 = 130.8 Hz to B7 = 3951.1 Hz) only one semitone narrower than used in Chapters 2–3. This result further suggests that two separate processes may allow extremely high pitches to be perceived as fast or slow depending on yet-to-be-determined factors.

In Chapter 5, I extended the analysis and discussion of the pitch discrimination experiment from Chapter 4. I found that bias-conforming responses tended to have substantially faster reaction times than bias-opposing responses and responses to correctly timed "bias–neutral" control tones. I also identified a possible negative correlation between pitch sensitivity and the strength of timing-induced bias. To test this correlation further, I presented a follow-up study measuring the effects of timing on perceived pitch under conditions of low and high pitch discriminability. However, I found similar levels of timing-induced bias under both difficulty conditions, as well as across individuals with different levels of pitch discrimination ability. I concluded by introducing Bayesian cue integration as a possible explanation for bidirectional influences between pitch and timing in auditory perception.

6.2 Theoretical Implications and Open Questions

The primary theoretical contributions of this thesis are therefore threefold:

First, I identified a novel negative quadratic relation between absolute pitch and perceived tempo, in which peak perceived tempo falls somewhere between A4 (440 Hz) and A6 (1760 Hz). This negative quadratic relation most likely has separate origins from the classic positive linear "higher–faster" effect of pitch on perceived tempo, although additional research will be needed to conclusively demonstrate their separability. Below I propose several such avenues for doing so.

Second, I established that higher–faster associations are bidirectional, in that the pitch of a sound influences its perceived timing and the timing of a sound influences its perceived pitch. This bidirectional influence provides evidence that pitch and timing are integrated into a unified representation, as hypothesized by several authors (Boltz, 2017; Henry & McAuley, 2009; Jones, 1976).

Third, my sensorimotor data suggest that the novel quadratic effect of pitch on timing arises upstream of the motor system. Whether the linear effect of pitch also influences motor timing within certain contexts remains an open question, as the present experiments found no evidence of such an influence.

6.2.1 Separate Origins of Linear and Quadratic Effects?

The primary question raised by the data I have presented is whether the quadratic effect of pitch that I observed in six out of eight perceived timing experiments has the same origin as the linear higher–faster illusion I observed in the remaining two (Chapters 2.5 & 4.3). I argued in Chapter 3 that the two effects most likely have separate origins, based on the fact that a change in pitch context produced a switch between quadratic and linear effects of pitch on subjective tempo, but not on motor timing. However, my pitch discrimination results from Chapters 4–5 also support the hypothesis that these are two separate effects.

Specifically, although I found evidence that the linear effect is bidirectional, I did not find similar evidence for the quadratic effect. In the pitch discrimination experiment that comprised Chapters 4.4 and 5.3, I tested the effect of tempo change on perceived pitch change at both A3 (220 Hz) and A5 (880 Hz). If the process that produces the quadratic effect from Chapters 2–3 also determines perceived pitch as a function of tempo, then we should expect the magnitude and directionality of timing-induced bias to differ between octaves. At A3, higher pitch should correspond to faster expected timing; therefore, earlier timing should produce higher perceived pitch, consistent with my findings. However, A5 is near the peak of the inverted U, meaning that perceived pitch and timing should be weakly correlated—or possibly even negatively correlated—at this octave. Instead, I found a nearly identical effect of timing on perceived pitch in both octaves (Figure 5.2), which is consistent with a simple linear association between timing and pitch.

From a theoretical standpoint, a quadratic relation between two variables is not invertible, which would complicate any bidirectional integration of pitch and time.

Although any given pitch would predict a single fast or slow tempo, any given tempo would predict two different levels of pitch—one high and one low. Strictly speaking, this does not rule out a learned relation between pitch and timing, as Bayesian priors need not be unimodal (McElreath, 2020) and the brain could still theoretically learn the relation. However, the lack of invertibility would mean that the polarity of the bidirectional relation between pitch changes and tempo changes would depend on the octave of the surrounding context, for which we currently have no evidence.

It is also more likely that the real-world pitch–timing correlation is linear, as there is little reason to expect most of the proposed origins discussed in Chapter 1.4 to be nonmonotonic. For example, correlations between size, pitch, and speed come from monotonic relations in physics, such as larger objects producing longer wavelengths of sound and requiring more energy to move. Similarly, high arousal correlates with faster speech due to both being higher-energy states, and there is little reason to expect increasing arousal to eventually produce slower speech, unless a point is reached where arousal disrupts communication. The acoustic explanation offered by Broze and Huron (2013) that faster attack transients allow higher pitches to be played faster is also based on physical laws that would not invert at extremely high pitches. As a counterpoint, however, their kinematic argument that greater motor demands cause low instruments to be played slowly might reasonably apply to extremely high pitches, at least for vocal production and certain instruments (e.g., playing at the highest positions on a violin). To conclusively demonstrate the shape of real-world pitch–timing correlations will require future analyses of cross-cultural speech and music datasets (Mehr et al., 2019; Ozaki et al., 2024).

6.2.2 Learned or Innate Effects?

As the real-world correlation between pitch and timing is most likely a monotonic higher–faster relation, the bidirectional, linear interactions between perceived pitch and timing might plausibly derive from a learned statistical prior. In contrast, the quadratic effect is unlikely to reflect the structure of the environment, and is therefore less likely to be a learned association. Consequently, the linear and quadratic effects of pitch on perceived timing may be differentiable by whether they are learned or innate.

Thus far, no illusory tempo research has been conducted in children or infants. The most relevant prior research has been in imagined motion, in which adults have been found to associate higher pitch with faster motion (Eitan & Granot, 2006), whereas children aged 5–11 have not (Eitan & Tubul, 2010; Kohn & Eitan, 2009, 2016). Tempo discrimination tasks have been conducted in children as young as 3 years of age (Bobin-Bègue & Provasi, 2005), and could be used to evaluate the timeline over which illusory tempo effects develop. To test whether pitch influences tempo perception in infants, however, a different approach would be needed. One promising avenue would be to use electroencephalography to test for neural correlates of illusory tempo. In particular, Herrmann et al. (2013) identified a tendency for the phase lag of stimulus-entrained oscillations in the auditory cortex to shift in opposite directions depending on whether pitch ascends or descends. Infant neural phase alignment to both drum rhythms and speech syllables has been found as early as 2 months of age (Ní Choisdealbha et al., 2023). Therefore, one could potentially apply pitch glides to these stimuli and assess whether pitch distorts infant neural tracking in the same manner as adults.

In adults, one could also test whether pitch–timing integration is malleable through implicit learning. In a pre-exposure phase, participants might complete either the temporal or pitch discrimination task from Chapter 4. Then, a set of melodies could be presented from an artificial musical grammar (e.g., Rohrmeier et al., 2011) in which pitch and tempo are negatively correlated. To increase the novelty and chance of learning, the pitches and melodies used could be derived from an unfamiliar musical system, such as the Bohlen-Pierce scale (Loui & Wessel, 2008; Loui et al., 2010; Loui, 2012). Finally, in a post-exposure phase, participants would repeat the task from the pre-exposure phase. If the bidirectional linear effect observed in Chapter 4 reflects a learned statistical relation, then the interactions between pitch and timing should weaken or reverse polarity after exposure.

In contrast, the quadratic effect of pitch on perceived tempo may be innate, perhaps rooted in the biophysics of the auditory periphery, similar to explanations for low-frequency timing superiority (Hove et al., 2014; Zuk et al., 2018) and high-frequency superiority for pitch perception (Trainor, Marie, Bruce, & Bidelman, 2014). The illusory tempo curve I observed in Chapters 2–3 bears a striking resemblance both to the pitch salience curve, which peaks at 700 Hz (Terhardt et al., 1982), and to simulated sub-cortical neural synchronization strength, which is weakest for sounds with approximately the same frequency (Zuk et al., 2018). Although it is currently unclear whether any of these three effects are related, it is conceivable that the ease of processing more salient sounds would speed up perceived timing (Eagleman, 2008; Eagleman & Pariyadath, 2009). I originally abandoned this idea after the final tempo rating experiment produced context-dependent effects; however, this hypothesis is worth revisiting if the quadratic and linear effects are indeed independent. One could attempt to control for differences in salience across octaves and potentially eliminate the quadratic effect by introducing noise to tones from each octave, inversely proportional to the pitch salience curve. It is also possible, as discussed in Chapter 3, that weaker neural synchronization combined with a general bias towards speeding up could explain why moderately high pitches are perceived as fastest.

6.2.3 Does Illusory Tempo Depend on Pitch or Frequency?

If the quadratic illusory tempo effect does relate to these similar innately-driven U-shaped effects, then we might expect it to depend not on pitch, but rather on the frequency content of a sound. Indeed, this may be one way to further disambiguate the linear and quadratic effects. If the quadratic effect arises due to the biophysics of the auditory system, but the linear effect arises at a later cue integration stage, then the former may depend on frequency while the latter depends on pitch. Thus far, no studies have attempted to disambiguate the illusory effects of frequency versus pitch, although Boltz (2011) did find an effect of timbral brightness on perceived tempo. Specifically, a timbre that had more energy at high-frequency harmonics was perceived as faster than one with more energy at lower harmonics. I also observed in the fourth experiment of Chapter 2 that synthetic complex tones were perceived as faster than piano tones of the same pitch, although I did not directly compare the brightness of the two timbres. Both of these findings are consistent with the frequency content of a sound influencing perceived tempo; however, neither experiment attempted to systematically separate the effects of pitch and frequency.

One method of testing dependence on frequency versus pitch would be to measure the perceived timing of complex tones with missing fundamental frequencies ($f0$), compared to pure tones of the same pitch—both in the relative tempo task from Chapter 2 and the temporal discrimination experiment from Chapter 4. For example, both a 220 Hz pure tone and a complex tone consisting of 440 Hz ($f1$), 660 Hz ($f2$), and 880 Hz ($f3$) sine waves can be perceived as having the same pitch: A3. If the quadratic effect of pitch actually depends on frequency, then the illusory tempo curve for pure tones should peak at least one octave higher than the curve for complex tones with missing fundamentals (as the lowest overtone is one octave above $f0$). If the quadratic effect instead depends on pitch, then peak perceived tempo should be at the same octave for both types of tones. If the linear effect also depends on frequency and not pitch, then missing-fundamental tones in the Chapter 4 temporal discrimination task should be perceived as being at least as early as pure tones that are one octave higher. However, if the linear effect depends on pitch as I hypothesized, then missing-fundamental tones should be perceived as similarly early as pure tones of the same pitch.

Together, these two tests might inform future modeling work by 1) offering insight into the stage of processing at which illusory tempo arises and 2) distinguishing whether there may be separate linear and quadratic effects of pitch/frequency on rhythm perception that originate from different stages of processing. Furthermore, differentiating effects of pitch versus frequency has real-world relevance, as the missing fundamental phenomenon is commonly used as a means of producing low-pitched sounds from speakers that are too small to generate the associated fundamental frequency. This means that a frequency-based effect might cause the same music or speech to produce different perceived timing when played over different audio devices.

6.2.4 Which Effect Dominates?

The factors that determine which effect dominates remain an open question, with the data from Chapters 2–3 providing a useful starting point. In the first four tempo rating experiments of Chapter 2, I generally observed asymmetrical inverted U-shaped curves, consistent with both a stronger negative quadratic effect and weaker positive linear effect of pitch on perceived tempo. In the final experiment of Chapter 2, I seemingly eliminated the negative quadratic effect while strengthening the positive linear effect by restricting the range of octaves present in the context from five to two and a half. In Chapter 3, however, I observed relatively symmetrical quadratic effects of pitch on motor timing, with the strength of that effect seemingly unaffected by the number of octaves present in the context.

One possibility is that the between-trial pitch context was more cohesive or salient in the experiments that showed linear effects than in those that showed quadratic effects. In the final tempo rating experiment of Chapter 2, the use of multiple notes per octave and the elimination of several-octave jumps between trials may have produced a more coherent pitch context by condensing pitch spacing, compared to the first four experiments. Similarly, the temporal discrimination task from Chapter 4.3 sampled eight pitch classes per octave rather than one, even though the overall pitch range tested was only one semitone less than used in Chapter 2. Alternatively it is possible that the use of only a single chroma (A, presented in different octaves) in some experiments reduced the salience of pitch changes between trials, dampening the use of pitch as a timing cue and allowing bottom-up effects of frequency to dominate.

Two modified versions of the Chapter 2 tempo rating experiments could be conducted to disambiguate effects of pitch spacing versus chroma variability. In one version of the task, one could present a different pitch class from each of the six octaves tested, perhaps by spacing the pitch conditions 13 semitones apart rather than 12. In the second version of the task, one could present two or more pitch classes from each of the six octaves. If the former produces a quadratic effect of pitch but the latter produces a linear effect, then having intervals smaller than an octave may be required to trigger the linear effect. If both versions of the task produce a higher–faster bias, then it is likely that the use of only a single chroma dampened the linear effect in several of my studies. Previous experiments in the field may have only detected positive linear effects due to not testing above 1000 Hz (e.g., Collier & Hubbard, 1998), only comparing two registers (e.g., Boltz, 2011) or within a single octave (Boltz, 2017; Feldstein & Bond, 1981; Pfeuty & Peretz, 2010), or using naturalistic melodies (Boltz, 2011), scales (Collier & Hubbard, 1998), and speech (Boltz, 2017).

I did, however, find a quadratic rather than linear effect in sensorimotor timing, even when two unique pitch classes were presented in Chapter 3.4. It may be the case that the substantially longer trials in the sensorimotor experiments (18 tones) compared to the subjective tempo (5 tones + 5 clicks) and temporal discrimination (6 sounds + 2 silent beats) tasks prevented a similar coherent context from forming. Another possibility is that the positive linear effect arises from a process that does not influence sensorimotor synchrony, such as memory encoding and retrieval (Boltz, 2017), or decision making. However, this explanation fails to account for previous evidence that pitch direction influences inter-tap intervals (Boasson & Granot, 2012) and the phase of neural entrainment in the auditory cortex (Herrmann et al., 2013).

6.3 Implications for Models of Rhythm Perception

Although many open questions remain, some progress can be made in incorporating pitch–timing interactions into current theoretical frameworks for rhythm perception. I will focus on incorporating the learned, linear association between higher pitch and faster timing, as I hypothesized that the quadratic effect of pitch likely arises subcortically or in the auditory periphery (Zuk et al., 2018), at an earlier stage of processing than is generally targeted by rhythm perception models. The data from Chapters 2–3 strongly support the need for further modeling work to explain why frequencies above 2000 Hz produce similarly strong neural synchronization as frequencies below 250 Hz. However, this branch of modeling lies beyond the scope of the present work.

6.3.1 Predictive Coding

In Chapter 5, I discussed at some length the idea that bidirectional influences of pitch and timing might be well explained as Bayesian cue integration (Friedman et al., 2013; Genewein et al., 2015). The most substantial implication of my data is that any Bayesian or predictive coding model of rhythm perception should attempt to jointly estimate pitch and timing, using sensory evidence from each dimension as a cue for the other. To validate the concept of illusory tempo and illusory pitch as products of Bayesian inference, similar to other visual (Weiss et al., 2002) and tactile (Goldreich & Tong, 2013) illusions, it will be necessary to demonstrate that the weighting of pitch versus timing information is precision-weighted. Although there is some evidence that pitch–timing integration is strongest when both dimensions are similarly salient (Prince, Schmuckler, & Thompson, 2009; Prince, 2011), this dependence has not been directly tested for pitch and timing biases—only factors such as pitch accuracy and perceived tonality. As it was suggested in Chapter 5, future research should test the effect of spectral degradation (e.g., Baer & Moore, 1993) on the relative weighting of temporal versus spectral cues in temporal and pitch discrimination tasks. If cross-dimensional biases result from Bayesian inference, then pitch-induced illusory tempo should weaken with spectral degradation, whereas timing-induced illusory pitch should strengthen.

At present, the most thoroughly-developed Bayesian model of rhythm perception is Phase and Tempo Inference from Point Process Event Timing (PATIPPET; Cannon, 2021). PATIPPET continuously estimates the underlying phase and tempo of a rhythm from the onset times of auditory events, using the posterior probability distribution from one moment in time as the prior probability for the next. It does not take into account information about the pitch of each stimulus, but Cannon (2021) does provide equations for a generalization of PATIPPET that could be used to generate different timing estimates for different event types. Thus, the model as it currently exists could be extended to allow pitch to bias phase and tempo estimates, such that higher pitches act as evidence of earlier phases and faster tempos. To capture the full bidirectionality of pitch–timing integration, however, PATIPPET would need to be extended to simultaneously infer the underlying pitch trajectory of rhythmic sequences. With this added dimension, the model would continuously estimate at each point in time the most likely pitch of an event, were one to occur at that moment. Upon observing an event, sensory information about the spectral properties of the sound would combine with this prior to determine perceived pitch and update the inferred trajectory of the rhythm and melody. Timing-induced illusory pitch could then be explained as a phase-dependency in the pitch estimate, wherein pitch increases are expected at phases leading up to the beat and pitch decreases are expected at phases slightly lagging the beat. This formulation bears a striking resemblance to Jones' (1976) original conceptualization of pitch–timing expectancy paths. It is, in essence, a Bayesian reformulation of the idea that the brain integrates pitch and timing into a unified representational space, the structure of which reflects real-world lawful relations between dimensions.

6.3.2 Neural Resonance Theory

A learned correlation between higher pitch and faster timing would be relatively straightforward to incorporate into Neural Resonance Theory (NRT) models. The simplest explanation is that oscillators in a gradient frequency network become attuned through Hebbian learning (Tichko et al., 2022) to the pitches that they most often entrain to. Based on real-world correlations, this would likely result in higher-tempo oscillators developing stronger connections to higher pitches. As pitch changes, the distribution of activation entering the gradient frequency network would shift up or down along the tempo map, driving sets of oscillators with slightly faster or slower preferred tempos depending on pitch direction, as hypothesized by Herrmann et al. (2013). A dynamical systems analysis by Kim and Large (2015) supports the idea that recruiting oscillators with different preferred tempos would bias the relative phase of entrainment without a change in its tempo—consistent with past MEG findings (Herrmann et al., 2013). One early simulation by Large (2000) tested a pitch-sensitive network like the one proposed here, and found that it better predicted human synchronization than a model unbiased by pitch. I therefore advocate for the reintroduction of this pitch-weighting to NRT gradient frequency networks. If the suggested architecture is correct, however, it remains unclear how contextual factors such as large pitch range, large interval size, or lack of chroma variance would be capable of eliminating higher–faster biases (Chapters 2–3).

The proposed architecture also does not account for the influence of timing change on perceived pitch. However, it may still be possible to explain timing-induced illusory pitch using the NRT framework. Consider the hypothetical entrained oscillation of cortical activity in the top panel of Figure 6.1, based on the canonical NRT oscillator (Large & Snyder, 2009):

$$\dot{z} = z(\alpha + i2\pi f + \beta|z|^2) + cs(t) \tag{6.3.1}$$

where α and β are parameters that determine the categorical behavior of the oscillator (typically set to $\alpha = 1$ and $\beta = -1$ to produce spontaneous oscillation with amplitude 1), f is the natural frequency of the oscillator in Hz, c is the coupling to a stimulus time series $s(t)$, and z is a complex number representing the state of the oscillator. The real part of z (blue line) represents the activity of an excitatory subpopulation of neurons, while the imaginary part (red line) represents the activity of an inhibitory subpopulation (Large et al., 2015).

Now consider how perceived pitch relates to this entrained activity. Although I have been referring to the effect of timing deviations on perceived pitch as linear, the periodic nature of rhythm means that timing deviations themselves are perceived nonlinearly. For example, a note arriving late by 10% of the interonset interval will most likely be perceived as late; however, a note arriving late by 70% of the interonset interval may be perceived as 70% late or 30% early, depending on whether it becomes

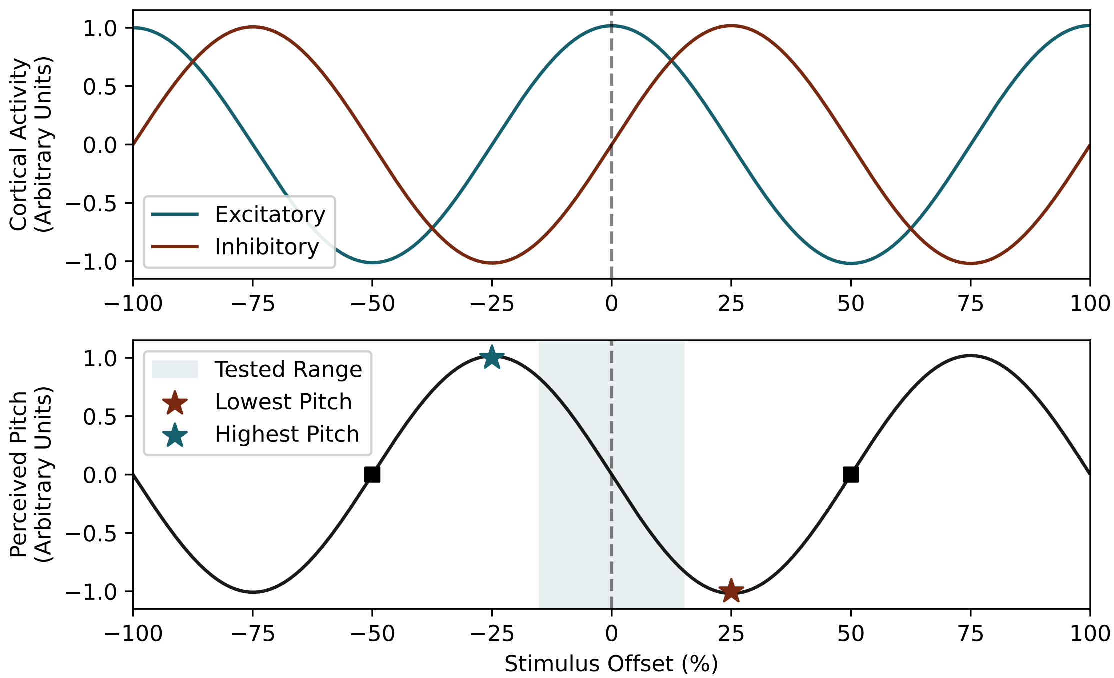

Figure 6.1: Hypothetical oscillations in perceived pitch during entrainment to an auditory rhythm. Top panel illustrates entrained oscillations of excitatory (blue) and inhibitory (red) subpopulations of neurons, as hypothesized by Neural Resonance Theory (Large et al., 2015; Large & Snyder, 2009). Bottom panel illustrates the hypothesized relation between stimulus phase and perceived pitch. Pitch is unbiased both on the beat and anti-phase of the beat (square markers). Perceived pitch peaks (blue star) when inhibitory activity is at its lowest and reaches its lowest (red star) when inhibitory activity is maximal.

perceptually associated with the beat preceding or following it. Therefore, the effect of deviations from rhythmic timing on perceived pitch is most likely only linear within a narrow window around the beat. In Chapters 4–5, I tested the perceived pitch of tones that were either 15% early, on the beat, or 15% late. I found that within this temporal region around the beat, later timing produced lower perceived pitch. This relation can be seen within the shaded region of the lower panel of Figure 6.1.

However, if we were to test a timing deviation of ±50% of the interonset interval, it would be ambiguous which beat the stimulus is tied to, and therefore whether it is early or late. Hypothetically, then, pitch should be unbiased at the anti-phase of the rhythm (square markers), as well as on the beat. Consequently, the greatest bias in perceived pitch should be halfway between the beat and its anti-phase, with stimuli being perceived as highest when 25% early (blue star) and lowest when 25% late (red star), before beginning to reverse direction as they approach the anti-phase. Extrapolating this relation gives the negative sine wave in the lower panel of Figure 6.1.

Notably, this hypothesis implies that the bias in perceived pitch is proportional to the derivative of the entrained oscillation, which itself is negatively proportional to the inhibitory population's activity. One possibility, then, is that the fluctuations in inhibitory activity during entrainment distorts perceived pitch. The lowest perceived pitch occurs when inhibitory activity peaks, and the highest perceived pitch occurs when inhibitory activity is weakest. Alternatively, pitch may be biased by the direction the entrained oscillation's tempo (or phase) is driven by the sound. In Dynamic Attending Theory (Large & Jones, 1999), Neural Resonance Theory (Large, 2010; Large et al., 2015; Large & Snyder, 2009), and related models (e.g., Adaptive Synchronization with Hebbian Learning and Elasticity; Roman et al., 2023), the tempo and/or phase of entrained oscillations are driven by stimulus phase via a negative sinusoidal relation that aligns with the bias in perceived pitch. Maximum perceived pitch occurs at the same stimulus phase that maximally drives the oscillator to speed up, and minimum perceived pitch occurs at the phase that maximally slows the oscillator. If a signal related to the change in the oscillator's timing was subsequently used as a cue for inferring pitch, it could produce this relation.

Both of these hypotheses depend on the assumption that perceived pitch is sinusoidally related to stimulus phase, which has yet to be empirically validated. One method for testing this hypothesis would be to modify the pitch discrimination task from Chapters 4–5 to use a long sequence of isochronous standard tones, within which pitch-shifted probe tones would fall at random phases. If my hypothesis is correct, the probability of rating tones as high-pitched should oscillate according to the negative sine of the probe's phase.

Lastly, it should be noted that the proposed NRT-based explanations for bidirectional influences of pitch and timing posit separate origins of higher pitches being perceived as faster (due to Hebbian learning) and faster timing being perceived as higher pitch (either due to inhibitory activity or a tempo change/error signal). This is in contrast to the predictive coding explanation that suggests a single Bayesian cue integration process that jointly estimates pitch and tempo/phase. In summary, my findings offer potential avenues for both leading theoretical frameworks of rhythm perception to incorporate pitch–timing integration.

6.4 Practical Applications for Auditory Interfaces

Although the goal of this thesis was to answer several theoretical questions that might bring us closer to an integrated model of pitch and rhythm perception, my results also have important practical applications for auditory interface design. Auditory interfaces have a broad range of medical, industrial, navigational, and military applications. They can be used in assistive devices for blindness by communicating spatial and graphical information non-visually (Edwards, 1989; Mynatt, 1997; R. D. Stevens, 1996), as well as to provide biofeedback for monitoring anxiety (Cheung, Han, Kushki, Anagnostou, & Biddiss, 2016). Auditory interfaces also help to distribute cognitive load across senses, and are therefore critical in settings where continual visual attention is required, such as while driving (Bazilinskyy & de Winter, 2015; Sodnik, Dicke, Tomažič, & Billinghurst, 2008), piloting aircraft, monitoring power stations (Sanderson, Anderson, & Watson, 2000), or playing video games (Coleman, Hand, Macaulay, & Newell, 2005).

In healthcare settings, auditory interfaces and alarms similarly allow doctors and nurses to monitor readings from multiple medical devices simultaneously, without looking away from the patient during treatment. Most often, these interfaces signal information using short three-tone sequences, with different pitch sequences corresponding to different devices and alerts (Gillard & Schutz, 2016). Unfortunately, the design of medical alarms and the complexity of the hospital environment often produce mistakes in alarm recognition (Edworthy, Hellier, Titchener, Naweed, & Roels, 2011; Gillard & Schutz, 2016). Recent efforts have been made to apply principles from music perception and auditory scene analysis (Bregman, 1990) to the design of medical interfaces and alarms (Foley, Anderson, & Schutz, 2020; Foley & Schutz, 2021). In particular, it has been noted that alarms with similar contours are often confusable (Gillard & Schutz, 2016), and their temporal and spectral properties seldom match sound structure that can be found in nature (Foley & Schutz, 2021). Improving the distinctiveness and learnability of alarm patterns will be critical for reducing mistakes and improving patient outcomes (Edworthy et al., 2011).

The pitch–timing biases identified in this thesis may provide one such avenue for improving the distinctiveness of auditory alarms. Introducing temporal variation in the interonset intervals of alarm tones could be used as a way to "perceptually highlight" the pitch contour. For example, if a tone sequence speeding up produces illusions of pitch increase, an ascending three-tone alarm might be made more distinctive or easier to learn by also making the second interonset interval shorter than the first. In contrast, an ascending alarm where the second interval was longer than the first might be less distinctive or more difficult to learn. In addition, auditory *tau* effects bias the perceived pitch of the middle tone in three-tone patterns towards the pitch of the temporally closer tone (Cohen et al., 1954). Varying interonset spacing could therefore also be used to make pitch intervals seem larger or smaller than they truly are, which could be exploited for improving the distinctiveness of alarm patterns. Future research should test the usefulness of timing-induced illusory pitch for improving the learnability and reducing the confusability of auditory alarms, perhaps in a learning and recognition task similar to that of Gillard and Schutz (2016). Previous work has already advocated for more natural sound design in the creation of alarm tones (Foley & Schutz, 2021). Perhaps we should introduce more natural pitch–timing correlations as well, given our brains' existing attunement to these structural properties.

6.5 Concluding Remarks

Rhythm perception is central to how we, as humans, interact with the world and with one another. Through an extensive program of research, this thesis has advanced our understanding of how pitch influences rhythm perception, and vice versa. I have demonstrated a novel and likely innate, negative quadratic relation between absolute pitch and both perceived tempo and sensorimotor timing. I have also demonstrated that the classic linear bias to perceive higher pitches as faster is, in fact, bidirectional. Although many questions remain open for future investigation, my data have offered a starting point for incorporating pitch into both predictive coding and neural resonance models of rhythm perception. Furthermore, my discovery that tempo change influences perceived pitch provides a promising new avenue for improving auditory interface design for medical alarms. It is my hope that we may begin integrating our theories of auditory perception, while improving patient outcomes, by understanding these bidirectional influences of pitch and time.

References

Addington, D. W. (1968). The relationship of selected vocal characteristics to personality perception. *Speech Monographs*, *35*(4), 492–503. doi: 10.1080/03637756809375599

Ammirante, P., & Thompson, W. F. (2010). Melodic accent as an emergent property of tonal motion. *Empirical Musicology Review*, *5*(3), 94–107. doi: 10.18061/1811/47559

Ammirante, P., & Thompson, W. F. (2012). Continuation tapping to triggered melodies: Motor resonance effects of melodic motion. *Experimental Brain Research*, *216*(1), 51–60. doi: 10.1007/s00221-011-2907-5

Ammirante, P., Thompson, W. F., & Russo, F. A. (2011). Ideomotor effects of pitch on continuation tapping. *Quarterly Journal of Experimental Psychology*, *64*(2), 381–393. doi: 10.1080/17470218.2010.495408

Apple, W., Streeter, L. A., & Krauss, R. M. (1979). Effects of pitch and speech rate and on personal attributions. *Journal of Personality and Social Psychology*, *37*(5), 715–727.

Baer, T., & Moore, B. C. J. (1993). Effects of spectral smearing on the intelligibility of sentences in noise. *The Journal of the Acoustical Society of America*, *94*(3), 1229–1241. doi: 10.1121/1.408176

Balkwill, L.-L., & Thompson, W. F. (1999). A cross-cultural investigation of the perception of emotion in music: Psychophysical and cultural cues. *Music Perception*, *17*(1), 43–64. doi: 10.2307/40285811

Battcock, A., & Schutz, M. (2019). Acoustically expressing affect. *Music Perception*, *37*(1), 66–91. doi: 10.1525/mp.2019.37.1.66

Bazilinskyy, P., & de Winter, J. (2015). Auditory interfaces in automated driving: an international survey. *PeerJ Computer Science*, *1*, e13. doi: 10.7717/peerj-cs.13

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

Bharucha, J. J. (1987). Music cognition and perceptual facilitation: A connectionist framework. *Music Perception*, *5*(1), 1–30. doi: 10.2307/40285384

Black, J. W. (1961). Relationships among fundamental frequency, vocal sound pressure, and rate of speaking. *Language and Speech*, *4*(4), 196–199. doi: 10.1177/002383096100400402

Boasson, A., & Granot, R. (2012). Melodic direction's effect on tapping. In *Proceedings of the 12th International Conference on Music Perception and Cognition and the 8th Triennial Conference of the European Society for the Cognitive Sciences of Music* (pp. 110–119). Thessaloniki, Greece.

Boasson, A., & Granot, R. (2019). Short latency effects of auditory frequency change on human motor behavior. *Auditory Perception & Cognition*, *2*(1–2), 98–128. doi: 10.1080/25742442.2019.1698264

Bobin-Bègue, A., & Provasi, J. (2005). Tempo discrimination in 3- and 4-year-old children: Performances and threshold. *Current psychology letters*(16, Vol. 2, 2005). doi: 10.4000/cpl.440

Boltz, M. G. (1998). Tempo discrimination of musical patterns: Effects due to pitch and rhythmic structure. *Perception & Psychophysics*, *60*(8), 1357–1373. doi: 10.3758/bf03207998

Boltz, M. G. (1999). The processing of melodic and temporal information: Independent or unified dimensions? *Journal of New Music Research*, *28*(1), 67–79. doi: 10.1076/jnmr.28.1.67.3121

Boltz, M. G. (2011). Illusory tempo changes due to musical characteristics. *Music Perception*, *28*(4), 367–386. doi: 10.1525/mp.2011.28.4.367

Boltz, M. G. (2017). Memory for vocal tempo and pitch. *Memory*, *25*(10), 1309–1326. doi: 10.1080/09658211.2017.1298808

Bose, A., Byrne, Á., & Rinzel, J. (2019). A neuromechanistic model for rhythmic beat generation. *PLOS Computational Biology*, *15*(5), e1006450. doi: 10.1371/journal.pcbi.1006450

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. The MIT Press. doi: 10.7551/mitpress/1486.001.0001

Brigner, W. L. (1988). Perceived duration as a function of pitch. *Perceptual and Motor Skills*, *67*(1), 301–302. doi: 10.2466/pms.1988.67.1.301

Brown, B. L., Strong, W. J., & Rencher, A. C. (1973). Perceptions of personality from speech: Effects of manipulations of acoustical parameters. *The Journal of the Acoustical Society of America*, *54*(1), 29–35. doi: 10.1121/1.1913571

Brown, S., & Jordania, J. (2011). Universals in the world's musics. *Psychology of Music*, *41*(2), 229–248. doi: 10.1177/0305735611425896

Broze, Y., & Huron, D. (2013). Is higher music faster? Pitch–speed relationships in Western compositions. *Music Perception*, *31*(1), 19–31. doi: 10.1525/mp.2013

.31.1.19

Butler, B. E., & Trainor, L. J. (2015). The musician redefined: A behavioral assessment of rhythm perception in professional club DJs. *Timing & Time Perception*, *3*(1-2), 116–132. doi: 10.1163/22134468-03002041

Cameron, D. J., Dotov, D., Flaten, E., Bosnyak, D., Hove, M. J., & Trainor, L. J. (2022). Undetectable very-low frequency sound increases dancing at a live concert. *Current Biology*, *32*(21), R1222–R1223. doi: 10.1016/j.cub.2022.09.035

Cannon, J. (2021). Expectancy-based rhythmic entrainment as continuous Bayesian inference. *PLOS Computational Biology*, *17*(6), e1009025. doi: 10.1371/journal.pcbi.1009025

Cason, N., & Schön, D. (2012). Rhythmic priming enhances the phonological processing of speech. *Neuropsychologia*, *50*(11), 2652–2658. doi: 10.1016/j.neuropsychologia.2012.07.018

Chang, A., Bosnyak, D. J., & Trainor, L. J. (2019). Rhythmicity facilitates pitch discrimination: Differential roles of low and high frequency neural oscillations. *NeuroImage*, *198*, 31–43. doi: 10.1016/j.neuroimage.2019.05.007

Cheung, S., Han, E., Kushki, A., Anagnostou, E., & Biddiss, E. (2016). Biomusic: An auditory interface for detecting physiological indicators of anxiety in children. *Frontiers in Neuroscience*, *10*. doi: 10.3389/fnins.2016.00401

Cirelli, L. K. (2018). How interpersonal synchrony facilitates early prosocial behavior. *Current Opinion in Psychology*, *20*, 35–39. doi: 10.1016/j.copsyc.2017.08.009

Cirelli, L. K., Einarson, K. M., & Trainor, L. J. (2014). Interpersonal synchrony increases prosocial behavior in infants. *Developmental Science*, *17*(6), 1003–1011. doi: 10.1111/desc.12193

Cirelli, L. K., Trehub, S. E., & Trainor, L. J. (2018). Rhythm and melody as social signals for infants. *Annals of the New York Academy of Sciences*, *1423*(1), 66–72. doi: 10.1111/nyas.13580

Cirelli, L. K., Wan, S. J., Johanis, T. C., & Trainor, L. J. (2018). Infants' use of interpersonal asynchrony as a signal for third-party affiliation. *Music & Science*, *1*, 205920431774585. doi: 10.1177/2059204317745855

Cohen, J., Hansel, C. E. M., & Sylvester, J. D. (1953). A new phenomenon in time judgment. *Nature*, *172*(4385), 901–901. doi: 10.1038/172901a0

Cohen, J., Hansel, C. E. M., & Sylvester, J. D. (1954). Interdependence of temporal and auditory judgments. *Nature*, *174*(4431), 642–644. doi: 10.1038/174642a0

Coleman, G. W., Hand, C., Macaulay, C., & Newell, A. F. (2005). Approaches to auditory interface design – lessons from computer games. In *Proceedings of ICAD 05-eleventh meeting of the international conference on auditory display* (pp. 99–104). Limerick, Ireland.

Collier, W. G., & Hubbard, T. L. (1998). Judgments of happiness, brightness, speed and tempo change of auditory stimuli varying in pitch and tempo. *Psychomusicology: A Journal of Research in Music Cognition*, *17*(1–2), 36–55. doi: 10.1037/h0094060

Collier, W. G., & Hubbard, T. L. (2001). Musical scales and evaluations of happiness and awkwardness: Effects of pitch, direction, and scale mode. *The American Journal of Psychology*, *114*(3), 355. doi: 10.2307/1423686

Cook, P., Rouse, A., Wilson, M., & Reichmuth, C. (2013). A California sea lion (Zalophus californianus) can keep the beat: Motor entrainment to rhythmic auditory stimuli in a non vocal mimic. *Journal of Comparative Psychology*, *127*(4), 412–427. doi: 10.1037/a0032345

Danielsen, A., Nymoen, K., Anderson, E., Câmara, G. S., Langerød, M. T., Thompson, M. R., & London, J. (2019). Where is the beat in that note? Effects of attack, duration, and frequency on the perceived timing of musical and quasi-musical sounds. *Journal of Experimental Psychology: Human Perception and Performance*, *45*(3), 402–418. doi: 10.1037/xhp0000611

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, *47*(1), 1–12. doi: 10.3758/s13428-014-0458-y

Drake, C., & Botte, M.-C. (1993). Tempo sensitivity in auditory sequences: Evidence for a multiple-look model. *Perception & Psychophysics*, *54*(3), 277–286. doi: 10.3758/bf03205262

Duke, R. A., Geringer, J. M., & Madsen, C. K. (1988). Effect of tempo on pitch perception. *Journal of Research in Music Education*, *36*(2), 108–125. doi: 10.2307/3345244

Dunovan, K. E., Tremel, J. J., & Wheeler, M. E. (2014). Prior probability and feature predictability interactively bias perceptual decisions. *Neuropsychologia*, *61*, 210–221. doi: 10.1016/j.neuropsychologia.2014.06.024

Eagleman, D. M. (2008). Human time perception and its illusions. *Current Opinion in Neurobiology*, *18*(2), 131–136. doi: 10.1016/j.conb.2008.06.002

Eagleman, D. M., & Pariyadath, V. (2009). Is subjective duration a signature of coding efficiency? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1525), 1841–1851. doi: 10.1098/rstb.2009.0026

Edwards, A. D. N. (1989). Soundtrack: An auditory interface for blind users. *Human–*

Computer Interaction, *4*(1), 45–66. doi: 10.1207/s15327051hci0401_2

Edworthy, J., Hellier, E., Titchener, K., Naweed, A., & Roels, R. (2011). Heterogeneity in auditory alarm sets makes them easier to learn. *International Journal of Industrial Ergonomics*, *41*(2), 136–146. doi: 10.1016/j.ergon.2010.12.004

Egger, S. W., Le, N. M., & Jazayeri, M. (2020). A neural circuit model for human sensorimotor timing. *Nature Communications*, *11*(1), 1–14. doi: 10.1038/s41467-020-16999-8

Eitan, Z., & Granot, R. Y. (2006). How music moves: Musical parameters and listeners' images of motion. *Music Perception*, *23*(3), 221–248. doi: 10.1525/mp.2006.23.3.221

Eitan, Z., & Tubul, N. (2010). Musical parameters and children's images of motion. *Musicae Scientiae*, *14*(2_suppl), 89–111. doi: 10.1177/10298649100140s207

Feldstein, S., & Bond, R. N. (1981). Perception of speech rate as a function of vocal intensity and frequency. *Language and Speech*, *24*(4), 387–394. doi: 10.1177/002383098102400408

Flach, R. (2005). The transition from synchronization to continuation tapping. *Human Movement Science*, *24*(4), 465–483. doi: 10.1016/j.humov.2005.09.005

Fletcher, H. (1924). The physical criterion for determining the pitch of a musical tone. *Physical Review*, *23*(3), 427–437. doi: 10.1103/physrev.23.427

Fletcher, H., & Munson, W. A. (1933). Loudness, its definition, measurement and calculation. *Bell System Technical Journal*, *12*(4), 377–430. doi: 10.1002/j.1538-7305.1933.tb00403.x

Foley, L., Anderson, C. J., & Schutz, M. (2020). Re-sounding alarms: Designing ergonomic auditory interfaces by embracing musical insights. *Healthcare*, *8*(4),

389. doi: 10.3390/healthcare8040389

Foley, L., & Schutz, M. (2021). High time for temporal variation: Improving sonic interaction with auditory interfaces. *IEEE Instrumentation & Measurement Magazine*, *24*(7), 4–9. doi: 10.1109/mim.2021.9549230

Friberg, A., & Sundberg, J. (1995). Time discrimination in a monotonic, isochronous sequence. *The Journal of the Acoustical Society of America*, *98*(5), 2524–2531. doi: 10.1121/1.413218

Friedman, A., Ludvig, E. A., Legge, E. L. G., & Vuong, Q. C. (2013). Bayesian combination of two-dimensional location estimates. *Behavior Research Methods*, *45*(1), 98–107. doi: 10.3758/s13428-012-0241-x

Friston, K. (2009). The free-energy principle: A rough guide to the brain? *Trends in Cognitive Sciences*, *13*(7), 293–301. doi: 10.1016/j.tics.2009.04.005

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138. doi: 10.1038/nrn2787

Friston, K., & Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1211–1221. doi: 10.1098/rstb.2008.0300

Gabrielsson, A., & Juslin, P. N. (1996). Emotional expression in music performance: Between the performer's intention and the listener's experience. *Psychology of Music*, *24*(1), 68–91. doi: 10.1177/0305735696241007

Genewein, T., Hez, E., Razzaghpanah, Z., & Braun, D. A. (2015). Structure learning in Bayesian sensorimotor integration. *PLOS Computational Biology*, *11*(8), e1004369. doi: 10.1371/journal.pcbi.1004369

Geringer, J. M., & Madsen, C. K. (1984). Pitch and tempo discrimination in recorded orchestral music among musicians and nonmusicians. *Journal of Research in Music Education*, *32*(3), 195–204. doi: 10.2307/3344838

Gillard, J., & Schutz, M. (2016). Composing alarms: considering the musical aspects of auditory alarm design. *Neurocase*, *22*(6), 566–576. doi: 10.1080/13554794.2016.1253751

Gioia, T. (2006). *Work Songs.* Durham, NC: Duke University Press. doi: 10.1515/9780822387688

Goldreich, D., & Tong, J. (2013). Prediction, postdiction, and perceptual length contraction: A Bayesian low-speed prior captures the cutaneous rabbit and related illusions. *Frontiers in Psychology*, *4*. doi: 10.3389/fpsyg.2013.00221

Gordon, M. S., & Ataucusi, A. (2021). Continuous sliding frequency shifts produce an illusory tempo drift. *JASA Express Letters*, *1*(5), 053202. doi: 10.1121/10.0005001

Gussenhoven, C., & Zhou, W. (2013). Revisiting pitch slope and height effects on perceived duration. In *Interspeech 2013.* ISCA. doi: 10.21437/interspeech.2013-360

Guyer, J. J., Fabrigar, L. R., & Vaughan-Johnston, T. I. (2018). Speech rate, intonation, and pitch: Investigating the bias and cue effects of vocal confidence on persuasion. *Personality and Social Psychology Bulletin*, *45*(3), 389–405. doi: 10.1177/0146167218787805

Gámez, J., Yc, K., Ayala, Y. A., Dotov, D., Prado, L., & Merchant, H. (2018). Predictive rhythmic tapping to isochronous and tempo changing metronomes in the nonhuman primate. *Annals of the New York Academy of Sciences*, *1423*(1), 396–414. doi: 10.1111/nyas.13671

Hautus, M. J. (1995). Corrections for extreme proportions and their biasing effects on estimated values of d'. *Behavior Research Methods, Instruments, & Computers*, *27*(1), 46–51. doi: 10.3758/bf03203619

Henry, M. J., & Herrmann, B. (2014). Low-frequency neural oscillations support dynamic attending in temporal context. *Timing & Time Perception*, *2*(1), 62–86. doi: 10.1163/22134468-00002011

Henry, M. J., Herrmann, B., & Obleser, J. (2014). Entrained neural oscillations in multiple frequency bands comodulate behavior. *Proceedings of the National Academy of Sciences*, *111*(41), 14935–14940. doi: 10.1073/pnas.1408741111

Henry, M. J., & McAuley, J. D. (2009). Evaluation of an imputed pitch velocity model of the auditory kappa effect. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(2), 551–564. doi: 10.1037/0096-1523.35.2.551

Henry, M. J., & McAuley, J. D. (2013). Perceptual distortions in pitch and time reveal active prediction and support for an auditory pitch-motion hypothesis. *PLoS ONE*, *8*(8), e70646. doi: 10.1371/journal.pone.0070646

Herbst, S. K., & Obleser, J. (2017). Implicit variations of temporal predictability: Shaping the neural oscillatory and behavioural response. *Neuropsychologia*, *101*, 141–152. doi: 10.1016/j.neuropsychologia.2017.05.019

Herbst, S. K., & Obleser, J. (2019). Implicit temporal predictability enhances pitch discrimination sensitivity and biases the phase of delta oscillations in auditory cortex. *NeuroImage*, *203*, 116198. doi: 10.1016/j.neuroimage.2019.116198

Herrmann, B., Henry, M. J., Grigutsch, M., & Obleser, J. (2013). Oscillatory phase dynamics in neural entrainment underpin illusory percepts of time. *Journal of Neuroscience*, *33*(40), 15799–15809. doi: 10.1523/jneurosci.1434-13.2013

Herrmann, B., Henry, M. J., Scharinger, M., & Obleser, J. (2014). Supplementary motor area activations predict individual differences in temporal-change sensitivity and its illusory distortions. *NeuroImage*, *101*, 370–379. doi: 10.1016/j.neuroimage.2014.07.026

Herrmann, B., & Johnsrude, I. S. (2018). Attentional state modulates the effect of an irrelevant stimulus dimension on perception. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(1), 89–105. doi: 10.1037/xhp0000432

Hove, M. J., Keller, P. E., & Krumhansl, C. L. (2007). Sensorimotor synchronization with chords containing tone-onset asynchronies. *Perception & Psychophysics*, *69*(5), 699–708. doi: 10.3758/bf03193772

Hove, M. J., Marie, C., Bruce, I. C., & Trainor, L. J. (2014). Superior time perception for lower musical pitch explains why bass-ranged instruments lay down musical rhythms. *Proceedings of the National Academy of Sciences*, *111*(28), 10383–10388. doi: 10.1073/pnas.1402039111

Hove, M. J., & Risen, J. L. (2009). It's all in the timing: Interpersonal synchrony increases affiliation. *Social Cognition*, *27*(6), 949–960. doi: 10.1521/soco.2009.27.6.949

Huron, D., & Parncutt, R. (1993). An improved model of tonality perception incorporating pitch salience and echoic memory. *Psychomusicology*, *12*(2), 154–171. doi: 10.1037/h0094110

International Organization for Standardization. (2003). *Acoustics—Normal equal-loudness-level contours (ISO 226:2003).*

International Organization for Standardization. (2017). *Acoustics—Methods for cal-*

culating loudness–Part 1: Zwicker method (ISO 532-1:2017).

International Telecommunication Union. (2017). *Recommendation ITU-R BS.1770-4 (10/2015): Algorithms to measure audio programme loudness and true-peak audio level.*

Jacoby, N., & McDermott, J. H. (2017). Integer ratio priors on musical rhythm revealed cross-culturally by iterated reproduction. *Current Biology*, *27*(3), 359–370. doi: 10.1016/j.cub.2016.12.031

Jacoby, N., Polak, R., Grahn, J. A., Cameron, D. J., Lee, K. M., Godoy, R., ... McDermott, J. H. (2024). Commonality and variation in mental representations of music revealed by a cross-cultural comparison of rhythm priors in 15 countries. *Nature Human Behaviour*. doi: 10.1038/s41562-023-01800-9

Jones, M. R. (1976). Time, our lost dimension: Toward a new theory of perception, attention, and memory. *Psychological Review*, *83*(5), 323–355. doi: 10.1037/0033-295x.83.5.323

Jones, M. R., & Boltz, M. (1989). Dynamic attending and responses to time. *Psychological Review*, *96*(3), 459–491. doi: 10.1037/0033-295x.96.3.459

Juslin, P. N. (1997). Emotional communication in music performance: A functionalist perspective and some data. *Music Perception*, *14*(4), 383–418. doi: 10.2307/40285731

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as Bayesian inference. *Annual Review of Psychology*, *55*(1), 271–304. doi: 10.1146/annurev.psych.55.090902.142005

Kim, J. C., & Large, E. W. (2015). Signal processing in periodically forced gradient frequency neural networks. *Frontiers in Computational Neuroscience*, *9*. doi:

10.3389/fncom.2015.00152

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends in Neurosciences*, *27*(12), 712–719. doi: 10.1016/j.tins.2004.10.007

Knill, D. C., & Richards, W. (Eds.). (1996). *Perception as Bayesian inference.* New York: Cambridge University Press.

Koelsch, S., Vuust, P., & Friston, K. (2019). Predictive processes and the peculiar case of music. *Trends in Cognitive Sciences*, *23*(1), 63–77. doi: 10.1016/j.tics.2018.10.006

Kohn, D., & Eitan, Z. (2009). Musical parameters and children's movement responses. In *7th Triennial Conference of European Society for the Cognitive Sciences of Music* (pp. 233–241). Jyväskylä, Finland.

Kohn, D., & Eitan, Z. (2016). Moving music: Correspondences of musical parameters and movement dimensions in children's motion and verbal responses. *Music Perception*, *34*(1), 40–55. doi: 10.1525/mp.2016.34.1.40

Korczynski, M. (2003). Music at work: Towards a historical overview. *Folk Music Journal*, *8*(3), 314–334.

Krumhansl, C. L., & Kessler, E. J. (1982). Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychological Review*, *89*(4), 334–368. doi: 10.1037/0033-295x.89.4.334

Kuhn, T. L., Wachhaus, G., Moore, R. S., & Pantle, J. E. (1979). Undergraduate nonmusic major vocal ranges. *Journal of Research in Music Education*, *27*(2), 68–75. doi: 10.2307/3344893

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package:

Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13). doi: 10.18637/jss.v082.i13

Lake, J. I., LaBar, K. S., & Meck, W. H. (2014). Hear it playing low and slow: How pitch level differentially influences time perception. *Acta Psychologica*, *149*, 169–177. doi: 10.1016/j.actpsy.2014.03.010

Large, E. W. (2000). On synchronizing movements to music. *Human Movement Science*, *19*(4), 527–566. doi: 10.1016/s0167-9457(00)00026-9

Large, E. W. (2010). Neurodynamics of music. In M. R. Jones, R. R. Fay, & A. N. Popper (Eds.), *Music perception* (Vol. 36, pp. 201–231). New York: Springer. doi: 10.1007/978-1-4419-6114-3

Large, E. W., Almonte, F. V., & Velasco, M. J. (2010). A canonical model for gradient frequency neural networks. *Physica D: Nonlinear Phenomena*, *239*(12), 905–911. doi: 10.1016/j.physd.2009.11.015

Large, E. W., & Gray, P. M. (2015). Spontaneous tempo and rhythmic entrainment in a bonobo (Pan paniscus). *Journal of Comparative Psychology*, *129*(4), 317–328. doi: 10.1037/com0000011

Large, E. W., Herrera, J. A., & Velasco, M. J. (2015). Neural networks for beat perception in musical rhythm. *Frontiers in Systems Neuroscience*, *9*, 159. doi: 10.3389/fnsys.2015.00159

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

Large, E. W., Kim, J. C., Flaig, N. K., Bharucha, J. J., & Krumhansl, C. L. (2016). A neurodynamic account of musical tonality. *Music Perception*, *33*(3), 319–331.

doi: 10.1525/mp.2016.33.3.319

Large, E. W., Roman, I., Kim, J. C., Cannon, J., Pazdera, J. K., Trainor, L. J., … Bose, A. (2023). Dynamic models for musical rhythm perception and coordination. *Frontiers in Computational Neuroscience*, *17*. doi: 10.3389/fncom.2023.1151895

Large, E. W., & Snyder, J. S. (2009). Pulse and meter as neural resonance. *Annals of the New York Academy of Sciences*, *1169*(1), 46–57. doi: 10.1111/j.1749-6632.2009.04550.x

Lenc, T., Keller, P. E., Varlet, M., & Nozaradan, S. (2018). Neural tracking of the musical beat is enhanced by low-frequency sounds. *Proceedings of the National Academy of Sciences*, *115*(32), 8221–8226. doi: 10.1073/pnas.1801421115

Liu, Y., Liu, G., Wei, D., Li, Q., Yuan, G., Wu, S., … Zhao, X. (2018). Effects of musical tempo on musicians' and non-musicians' emotional experience when listening to music. *Frontiers in Psychology*, *9*. doi: 10.3389/fpsyg.2018.02118

Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, *1*(4), 476–490. doi: 10.3758/bf03210951

London, J., Thompson, M., Burger, B., Hildreth, M., & Toiviainen, P. (2019). Tapping doesn't help: Synchronized self-motion and judgments of musical tempo. *Attention, Perception, & Psychophysics*, *81*(7), 2461–2472. doi: 10.3758/s13414-019-01722-7

Loui, P. (2012). Learning and liking of melody and harmony: Further studies in artificial grammar learning. *Topics in Cognitive Science*, *4*(4), 554–567. doi: 10.1111/j.1756-8765.2012.01208.x

Loui, P., & Wessel, D. (2008). Learning and liking an artificial musical system: effects of set size and repeated exposure. *Musicae Scientiae*, *12*(2), 207–230. doi: 10.1177/102986490801200202

Loui, P., Wessel, D. L., & Kam, C. L. H. (2010). Humans rapidly learn grammatical structure in a new musical scale. *Music Perception*, *27*(5), 377–388. doi: 10.1525/mp.2010.27.5.377

Madsen, C. K., Duke, R. A., & Geringer, J. M. (1984). Pitch and tempo discrimination in recorded band music among wind and percussion musicians. *Journal of Band Research*, *20*(1), 20–29.

Mairesse, F., & Walker, M. (2006). Automatic recognition of personality in conversation. In *Proceedings of the Human Language Technology Conference of the NAACL* (pp. 85–88). Association for Computational Linguistics. doi: 10.3115/1614049.1614071

Manning, F. C., & Schutz, M. (2013). "Moving to the beat" improves timing perception. *Psychonomic Bulletin & Review*, *20*(6), 1133–1139. doi: 10.3758/s13423-013-0439-7

Manning, F. C., & Schutz, M. (2016). Trained to keep a beat: Movement-related enhancements to timing perception in percussionists and non-percussionists. *Psychological Research*, *80*(4), 532–542. doi: 10.1007/s00426-015-0678-5

Mates, J. (1994a). A model of synchronization of motor acts to a stimulus sequence: II. Stability analysis, error estimation and simulations. *Biological Cybernetics*, *70*(5), 475–484. doi: 10.1007/bf00203240

Mates, J. (1994b). A model of synchronization of motor acts to a stimulus sequence: I. Timing and error corrections. *Biological Cybernetics*, *70*(5), 463–473. doi:

10.1007/bf00203239

McAuley, J. D., & Fromboluti, E. K. (2014). Attentional entrainment and perceived event duration. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1658), 20130401. doi: 10.1098/rstb.2013.0401

McAuley, J. D., Jones, M. R., Holub, S., Johnston, H. M., & Miller, N. S. (2006). The time of our lives: Life span development of timing and event tracking. *Journal of Experimental Psychology: General*, *135*(3), 348–367. doi: 10.1037/0096-3445.135.3.348

McElreath, R. (2020). *Statistical rethinking: A Bayesian course with examples in R and Stan* (2nd ed.). Chapman & Hall. doi: 10.1201/9780429029608

Mehr, S. A., Singh, M., Knox, D., Ketter, D. M., Pickens-Jones, D., Atwood, S., ... Glowacki, L. (2019). Universality and diversity in human song. *Science*, *366*(6468). doi: 10.1126/science.aax0868

Michon, J. A. (1967). *Timing in temporal tracking.* Soesterberg, Netherlands: Institute for Perception RVO-TNO.

Miller, N., Maruyama, G., Beaber, R. J., & Valone, K. (1976). Speed of speech and persuasion. *Journal of Personality and Social Psychology*, *34*(4), 615–624. doi: 10.1037/0022-3514.34.4.615

Moore, R. S. (1991). Comparison of children's and adults' vocal ranges and preferred tessituras in singing familiar songs. *Bulletin of the Council for Research in Music Education*, 13–22.

Mynatt, E. (1997). Transforming graphical interfaces into auditory interfaces for blind users. *Human-Computer Interaction*, *12*(1), 7–45. doi: 10.1207/s15327051hci1201&2_2

Neuhoff, J. G. (2016). Looming sounds are perceived as faster than receding sounds. *Cognitive Research: Principles and Implications*, *1*(1). doi: 10.1186/s41235-016-0017-4

Ní Choisdealbha, A., Attaheri, A., Rocha, S., Mead, N., Olawole-Scott, H., Brusini, P., ... Goswami, U. (2023). Neural phase angle from two months when tracking speech and non-speech rhythm linked to language performance from 12 to 24 months. *Brain and Language*, *243*, 105301. doi: 10.1016/j.bandl.2023.105301

Ozaki, Y., Tierney, A., Pfordresher, P. Q., McBride, J. M., Benetos, E., Proutskova, P., ... Savage, P. E. (2024). Globally, songs and instrumental melodies are slower and higher and use more stable pitches than speech: A Registered Report. *Science Advances*, *10*(20). doi: 10.1126/sciadv.adm9797

Palmer, C., & Demos, A. P. (2022). Are we in time? How predictive coding and dynamical systems explain musical synchrony. *Current Directions in Psychological Science*, *31*(2), 147–153. doi: 10.1177/09637214211053635

Patel, A. D., Iversen, J. R., Bregman, M. R., & Schulz, I. (2009a). Experimental evidence for synchronization to a musical beat in a nonhuman animal. *Current Biology*, *19*(10), 827–830. doi: 10.1016/j.cub.2009.03.038

Patel, A. D., Iversen, J. R., Bregman, M. R., & Schulz, I. (2009b). Studying synchronization to a musical beat in nonhuman animals. *Annals of the New York Academy of Sciences*, *1169*(1), 459–469. doi: 10.1111/j.1749-6632.2009.04581.x

Pazdera, J. K., Rinaldi, O. M., & Trainor, L. (2024). Timing-induced illusory percepts of pitch. *PsyArXiv*. doi: 10.31234/osf.io/caxsb

Pazdera, J. K., & Trainor, L. J. (2023). Bidirectional interactions of pitch and time. In *Proceedings of the 17th International Conference on Music Perception and Cognition* (pp. 301–306). Tokyo, Japan.

Pazdera, J. K., & Trainor, L. J. (2024a). Pitch biases sensorimotor synchronization to auditory rhythms. *PsyArXiv*. doi: 10.31234/osf.io/fbmw5

Pazdera, J. K., & Trainor, L. J. (2024b). Pitch-induced illusory percepts of time. *PsyArXiv*. doi: 10.31234/osf.io/6fx87

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., ... Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. doi: 10.3758/s13428-018-01193-y

Pfeuty, M., & Peretz, I. (2010). Abnormal pitch-time interference in congenital amusia: Evidence from an implicit test. *Attention, Perception, & Psychophysics*, *72*(3), 763–774. doi: 10.3758/app.72.3.763

Poppe, D. (2010). Scribing work songs at an archeological dig in Egypt. *General Music Today*, *24*(2), 31–36. doi: 10.1177/1048371310385214

Prince, J. B. (2011). The integration of stimulus dimensions in the perception of music. *Quarterly Journal of Experimental Psychology*, *64*(11), 2125–2152. doi: 10.1080/17470218.2011.573080

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. doi: 10.3758/mc.37.3.368

Prpic, V., Murgia, M., Tommaso, M. D., Boschetti, G., Galmonte, A., & Agostini, T. (2016). Octave bias in pitch perception: The influence of pitch height on pitch class identification. *Perception*, *45*(9), 1060–1069. doi: 10.1177/

0301006616651953

Ray, G. B. (1986). Vocally cued personality prototypes: An implicit personality theory approach. *Communication Monographs*, *53*(3), 266–276. doi: 10.1080/03637758609376141

Repp, B. H. (2003). Phase attraction in sensorimotor synchronization with auditory sequences: Effects of single and periodic distractors on synchronization accuracy. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(2), 290–309. doi: 10.1037/0096-1523.29.2.290

Repp, B. H. (2005). Sensorimotor synchronization: A review of the tapping literature. *Psychonomic Bulletin & Review*, *12*(6), 969–992. doi: 10.3758/bf03206433

Repp, B. H. (2006). Rate limits of sensorimotor synchronization. *Advances in Cognitive Psychology*, *2*(2), 163–181. doi: 10.2478/v10053-008-0053-9

Repp, B. H., & Su, Y.-H. (2013). Sensorimotor synchronization: A review of recent research (2006–2012). *Psychonomic Bulletin & Review*, *20*(3), 403–452. doi: 10.3758/s13423-012-0371-2

Rohrmeier, M., Rebuschat, P., & Cross, I. (2011). Incidental and online learning of melodic structure. *Consciousness and Cognition*, *20*(2), 214–222. doi: 10.1016/j.concog.2010.07.004

Roman, I. R., Roman, A. S., Kim, J. C., & Large, E. W. (2023). Hebbian learning with elasticity explains how the spontaneous motor tempo affects music performance synchronization. *PLOS Computational Biology*, *19*(6), e1011154. doi: 10.1371/journal.pcbi.1011154

Rothermich, K., Schmidt-Kassow, M., & Kotz, S. A. (2012). Rhythm's gonna get you: Regular meter facilitates semantic sentence processing. *Neuropsychologia*, *50* (2), 232–244. doi: 10.1016/j.neuropsychologia.2011.10.025

Ruggero, M. A. (1992). Responses to sound of the basilar membrane of the mammalian cochlea. *Current Opinion in Neurobiology*, *2* (5), 690. doi: 10.1016/0959-4388(92)90040-r

Sanderson, P. M., Anderson, J., & Watson, M. (2000). Extending ecological interface design to auditory displays. In *Proceedings of the 2000 Annual Conference of the Computer-Human Interaction Special Interest Group (CHISIG) of the Ergonomics Society of Australia (OzCHI2000)* (pp. 259–266). Sydney: CSIRO.

Sares, A. G., Foster, N. E. V., Allen, K., & Hyde, K. L. (2018). Pitch and time processing in speech and tones: The effects of musical training and attention. *Journal of Speech, Language, and Hearing Research*, *61* (3), 496–509. doi: 10.1044/2017_jslhr-s-17-0207

Scherer, K. R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin*, *99* (2), 143–165. doi: 10.1037/0033-2909.99.2.143

Scherer, K. R., & Oshinsky, J. S. (1977). Cue utilization in emotion attribution from auditory stimuli. *Motivation and Emotion*, *1* (4), 331–346. doi: 10.1007/bf00992539

Schultz, B. G., & van Vugt, F. T. (2016). Tap Arduino: An Arduino microcontroller for low-latency auditory feedback in sensorimotor synchronization experiments. *Behavior Research Methods*, *48* (4), 1591–1607. doi: 10.3758/s13428-015-0671-3

Schutz, M., & Vaisberg, J. M. (2012). Surveying the temporal structure of sounds used in *Music Perception*. *Music Perception*, *31* (3), 288–296. doi: 10.1525/

mp.2014.31.3.288

Smith, B. L., Brown, B. L., Strong, W. J., & Rencher, A. C. (1975). Effects of speech rate on personality perception. *Language and Speech*, *18*(2), 145–152. doi: 10.1177/002383097501800203

Sodnik, J., Dicke, C., Tomažič, S., & Billinghurst, M. (2008). A user study of auditory versus visual interfaces for use while driving. *International Journal of Human-Computer Studies*, *66*(5), 318–332. doi: 10.1016/j.ijhcs.2007.11.001

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, *31*(1), 137–149. doi: 10.3758/bf03207704

Steinmetz, C. (2019). pyloudnorm (Version v0.1.0) [Computer software manual].

Stevens, L. T. (1886). On the time-sense. *Mind*, *43*(11), 393–404.

Stevens, R. D. (1996). *Principles for the design of auditory interfaces to present complex information to blind people* (Unpublished doctoral dissertation). The University of York.

Stupacher, J., Hove, M. J., & Janata, P. (2016). Audio features underlying perceived groove and sensorimotor synchronization in music. *Music Perception*, *33*(5), 571–589. doi: 10.1525/mp.2016.33.5.571

Tal, I., Large, E. W., Rabinovitch, E., Wei, Y., Schroeder, C. E., Poeppel, D., & Golumbic, E. Z. (2017). Neural entrainment to the beat: The "missing-pulse" phenomenon. *The Journal of Neuroscience*, *37*(26), 6331–6341. doi: 10.1523/jneurosci.2500-16.2017

Tamir-Ostrover, H., & Eitan, Z. (2015). Higher is faster: Pitch register and tempo preferences. *Music Perception*, *33*(2), 179–198. doi: 10.1525/mp.2015.33.2.179

Terhardt, E., Stoll, G., & Seewann, M. (1982). Algorithm for extraction of pitch and pitch salience from complex tonal signals. *The Journal of the Acoustical Society of America*, *71*(3), 679–688. doi: 10.1121/1.387544

Tichko, P., Kim, J. C., & Large, E. W. (2022). A dynamical, radically embodied, and ecological theory of rhythm development. *Frontiers in Psychology*, *13*. doi: 10.3389/fpsyg.2022.653696

Tichko, P., & Large, E. W. (2019). Modeling infants' perceptual narrowing to musical rhythms: Neural oscillation and Hebbian plasticity. *Annals of the New York Academy of Sciences*, *1453*(1), 125–139. doi: 10.1111/nyas.14050

Trainor, L. J., Marie, C., Bruce, I. C., & Bidelman, G. M. (2014). Explaining the high voice superiority effect in polyphonic music: Evidence from cortical evoked potentials and peripheral auditory models. *Hearing Research*, *308*, 60–70. doi: 10.1016/j.heares.2013.07.014

Trainor, L. J., & Trehub, S. E. (1992). A comparison of infants' and adults' sensitivity to Western musical structure. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(2), 394–402. doi: 10.1037/0096-1523.18.2.394

Trainor, L. J., & Trehub, S. E. (1994). Key membership and implied harmony in Western tonal music: Developmental perspectives. *Perception & Psychophysics*, *56*(2), 125–132. doi: 10.3758/bf03213891

Trochidis, K., & Bigand, E. (2013). Investigation of the effect of mode and tempo on emotional responses to music using EEG power asymmetry. *Journal of Psychophysiology*, *27*(3), 142–148. doi: 10.1027/0269-8803/a000099

Törnqvist, L., Vartia, P., & Vartia, Y. O. (1985). How should relative changes be measured? *The American Statistician*, *39*(1), 43–46. doi: 10.1080/00031305

.1985.10479385

Varlet, M., Williams, R., & Keller, P. E. (2020). Effects of pitch and tempo of auditory rhythms on spontaneous movement entrainment and stabilisation. *Psychological Research*, *84*, 568–584. doi: 10.1007/s00426-018-1074-8

Vilares, I., & Kording, K. (2011). Bayesian models: The structure of the world, uncertainty, behavior, and the brain. *Annals of the New York Academy of Sciences*, *1224*(1), 22–39. doi: 10.1111/j.1749-6632.2011.05965.x

Vincent, B. T. (2015). A tutorial on Bayesian models of perception. *Journal of Mathematical Psychology*, *66*, 103–114. doi: 10.1016/j.jmp.2015.02.001

Vos, P. G., van Assen, M., & Fraňek, M. (1997). Perceived tempo change is dependent on base tempo and direction of change: Evidence for a generalized version of Schulze's (1978) internal beat model. *Psychological Research*, *59*(4), 240–247. doi: 10.1007/bf00439301

Vuust, P., Dietz, M. J., Witek, M., & Kringelbach, M. L. (2018). Now you hear it: A predictive coding model for understanding rhythmic incongruity. *Annals of the New York Academy of Sciences*, *1423*(1), 19–29. doi: 10.1111/nyas.13622

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

Walker, P., & Smith, S. (1984). Stroop interference based on the synaesthetic qualities of auditory pitch. *Perception*, *13*(1), 75–81. doi: 10.1068/p130075

Weiss, Y., Simoncelli, E. P., & Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nature Neuroscience*, *5*(6), 598–604. doi: 10.1038/nn0602-858

Wilson, M., & Cook, P. F. (2016). Rhythmic entrainment: Why humans want to, fireflies can't help it, pet birds try, and sea lions have to be bribed. *Psychonomic Bulletin & Review*, *23*(6), 1647–1659. doi: 10.3758/s13423-016-1013-x

Wiltermuth, S. S., & Heath, C. (2009). Synchrony and cooperation. *Psychological Science*, *20*(1), 1–5. doi: 10.1111/j.1467-9280.2008.02253.x

Wing, A., & Kristofferson, A. (1973). The timing of interresponse intervals. *Perception & Psychophysics*, *13*(3), 455–460. doi: 10.3758/bf03205802

Wojtczak, M., Mehta, A. H., & Oxenham, A. J. (2017). Rhythm judgments reveal a frequency asymmetry in the perception and neural coding of sound synchrony. *Proceedings of the National Academy of Sciences*, *114*(5), 1201–1206. doi: 10.1073/pnas.1615669114

Woods, K. J. P., Siegel, M. H., Traer, J., & McDermott, J. H. (2017). Headphone screening to facilitate web-based auditory experiments. *Attention, Perception, & Psychophysics*, *79*(7), 2064–2072. doi: 10.3758/s13414-017-1361-2

Zamm, A., Wang, Y., & Palmer, C. (2018). Musicians' natural frequencies of performance display optimal temporal stability. *Journal of Biological Rhythms*, *33*(4), 432–440. doi: 10.1177/0748730418783651

Zuk, N. J., Carney, L. H., & Lalor, E. C. (2018). Preferred tempo and low-audio-frequency bias emerge from simulated sub-cortical processing of sounds with a musical beat. *Frontiers in Neuroscience*, *12*. doi: 10.3389/fnins.2018.00349