A TEST OF AN AUDITORY MOTION HYPOTHESIS FOR CONTINUOUS AND DISCRETE SOUNDS MOVING IN PITCH SPACE

Molly J. Henry

A Dissertation
Submitted to the Graduate College of Bowling Green State University in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

May 2011

Committee:
J. Devin McAuley, Advisor
Rodney M. Gabel
Graduate Faculty Representative
Jennifer Z. Gillespie
Dale S. Klopfer
Ten experiments tested an auditory motion hypothesis, which proposes that regular pitch-time trajectories facilitate perception of and attention to auditory stimuli; on this view, listeners are assumed to use velocity information (pitch change per unit time) to generate expectations about the future time course of continuous and discrete sounds moving in pitch space. Toward this end, two sets of experiments were conducted. In six experiments reported in Part I of this dissertation, listeners judged the duration or pitch change of a continuous or discrete comparison stimulus relative to a standard, where the comparison’s velocity varied on each trial relative to the fixed standard velocity. Results indicate that expectations generated based on velocity information led to distortions in perceived duration and pitch change of continuous stimuli that were consistent with the auditory motion hypothesis; specifically, when comparison velocity was relatively fast, duration was overestimated and pitch change was underestimated. Moreover, when comparison velocity was relatively slow, duration was underestimated and pitch change was overestimated. On the other hand, no perceptual distortions were observed for discrete stimuli, consistent with the idea that velocity information is less clearly conveyed, or easier to ignore, for discrete auditory stimuli.

Four experiments reported in Part II tested the hypothesis that listeners tune attention to expected pitch-time locations of future events based on velocity information conveyed by continuous and discrete auditory stimuli. Listeners detected pure-tone signals in noise that were expected or unexpected based on extrapolation of the trajectory of ascending or descending glides or sequences. Consistent with the auditory motion hypothesis, results indicate that listeners
used pitch-time trajectory information in continuous and discrete auditory stimuli to tune attention; that is, listeners were most sensitive to detect expected relative to unexpected signals, and results were similar for continuous and discrete cues. However, consistent with an auditory gravity hypothesis, large asymmetries were observed for ascending versus descending cues, with listeners overshooting the expected pitch location of signals following descending cues. Taken together, the results support the hypothesis that listeners use velocity information to generate expectations about the future time course of sounds moving in pitch space.
ACKNOWLEDGMENTS

I have been lucky to have received an amazing amount of support through not only the completion of this manuscript, but in the years leading to up this point. It is my honor to be able to acknowledge those who have made my education a fun and fulfilling endeavor. First I must acknowledge the unconditional love and support that my family has given me all of my life. Without them, it is impossible to imagine myself doing what I am today. To my parents, Phil and Mary, and my brother, Paul: thank you and I love you. Also, to my cat Tweeker, who sat on my lap for a large portion of the time spent writing this dissertation, thanks.

A million thanks are due to my advisor, Devin McAuley. It was by accident that I ended up in his lab, and I cannot imagine a better outcome. Devin has provided me with a challenging and productive work environment for the past five years that I believe has allowed me to realize my potential as a scientist. I have immensely enjoyed my time working and learning with Devin, and I look forward to our future interactions.

Many thanks are also due to William Hartmann, professor of physics and head of the psychoacoustics lab at Michigan State University. Since I arrived at MSU just over a year ago, Bill has adopted me into his lab, allowed me use of his equipment and knowledge, and demonstrated infinite patience in his dealings with me. Bill was instrumental in completion of this manuscript, and for that he deserves many thanks. Thanks also to Tim Pleskac for much useful discussion regarding the simulations presented in this dissertation.

I would like to acknowledge the support of my labmates, present and past, who have shared their knowledge, logged hours of useful discussion, and participated in non-scholastic activities that nonetheless make the scholastic hours more bearable. Laura Dilley, Nate Miller, and Louis Vinke deserve special mention in this regard. Laura Dilley has given me the opportunity to learn many skills, and has provided me with unfailing support and sound advice.
Nate Miller has provided me with meaningful advice and friendship through all of my graduate training. As an example of this, during these final weeks of manuscript preparation, Nate has sent daily emails containing ‘dissertation motivators’ to keep morale high. Finally, Louis Vinke has been a formidable ‘partner in crime’ in many shared adventures.

Over the years, I have been helped by research assistants who all deserve mention, but are too numerous to name. Several that deserve special thanks are those who helped with data collection for the experiments contained in this dissertation. So thanks to Liza Kustantin, Ben Mastay, Brian Spyke, Neelima Wagley, Jon Walters, Alan Wedd, Amelia Wiggins, Phyllis Yan, and Lauren Zeilinski, I would also like to acknowledge the listeners who participated in Experiments 8, 9, and 10 of this dissertation, who each contributed a minimum of six hours to testing, mostly for an interest in science.

Finally, thanks are due to the members of my committee, Dale Klopfer, Jen Gillespie, and Rod Gabel. The three of you have provided useful comments that have helped to shape this dissertation.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td><strong>PART I</strong></td>
<td></td>
</tr>
<tr>
<td>CHAPTER 1. PERSPECTIVES ON THE PERCEPTUAL INTERDEPENDENCE OF PITCH AND TIME</td>
<td>6</td>
</tr>
<tr>
<td>Perceptual Independence of Pitch and Time</td>
<td>8</td>
</tr>
<tr>
<td>Behavioral evidence for pitch-time independence</td>
<td>8</td>
</tr>
<tr>
<td>Neural evidence for pitch-time independence</td>
<td>11</td>
</tr>
<tr>
<td>Acquired amusia (lesion studies)</td>
<td>11</td>
</tr>
<tr>
<td>Neuroimaging studies</td>
<td>13</td>
</tr>
<tr>
<td>Summary</td>
<td>16</td>
</tr>
<tr>
<td>Perceptual Interdependence of Pitch and Time</td>
<td>16</td>
</tr>
<tr>
<td>Behavioral evidence for pitch-time interdependence</td>
<td>17</td>
</tr>
<tr>
<td>Neural evidence for pitch-time interdependence</td>
<td>22</td>
</tr>
<tr>
<td>Auditory motion hypothesis</td>
<td>24</td>
</tr>
<tr>
<td><strong>CHAPTER 2: APPLICATION OF THE AUDITORY MOTION HYPOTHESIS</strong></td>
<td></td>
</tr>
<tr>
<td>TO SOUNDS MOVING CONTINUOUSLY IN PITCH SPACE</td>
<td>33</td>
</tr>
<tr>
<td>Experiment 1</td>
<td>36</td>
</tr>
<tr>
<td>Method</td>
<td>36</td>
</tr>
<tr>
<td>Design</td>
<td>36</td>
</tr>
<tr>
<td>Participants</td>
<td>36</td>
</tr>
<tr>
<td>Stimuli and apparatus</td>
<td>36</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Procedure</td>
<td>37</td>
</tr>
<tr>
<td>Data analysis</td>
<td>39</td>
</tr>
<tr>
<td>Results</td>
<td>41</td>
</tr>
<tr>
<td>Imputed velocity model</td>
<td>45</td>
</tr>
<tr>
<td>Summary</td>
<td>45</td>
</tr>
<tr>
<td>Experiment 2</td>
<td>46</td>
</tr>
<tr>
<td>Method</td>
<td>46</td>
</tr>
<tr>
<td>Design</td>
<td>46</td>
</tr>
<tr>
<td>Participants</td>
<td>46</td>
</tr>
<tr>
<td>Stimuli and apparatus</td>
<td>47</td>
</tr>
<tr>
<td>Procedure</td>
<td>47</td>
</tr>
<tr>
<td>Results</td>
<td>47</td>
</tr>
<tr>
<td>Imputed velocity model</td>
<td>51</td>
</tr>
<tr>
<td>Summary</td>
<td>52</td>
</tr>
<tr>
<td>Experiment 3</td>
<td>53</td>
</tr>
<tr>
<td>Method</td>
<td>53</td>
</tr>
<tr>
<td>Design</td>
<td>53</td>
</tr>
<tr>
<td>Participants</td>
<td>53</td>
</tr>
<tr>
<td>Stimuli and apparatus</td>
<td>53</td>
</tr>
<tr>
<td>Procedure</td>
<td>55</td>
</tr>
<tr>
<td>Results</td>
<td>55</td>
</tr>
<tr>
<td>Imputed velocity model</td>
<td>58</td>
</tr>
<tr>
<td>Comparison of results across standard velocities</td>
<td>58</td>
</tr>
</tbody>
</table>
CHAPTER 3: EXTENSION OF THE AUDITORY MOTION HYPOTHESIS TO SOUNDS IMPLYING MOTION IN FREQUENCY SPACE

Experiment 4

Method

Design

Participants

Stimuli and apparatus

Procedure

Results

Imputed velocity model

Summary

Experiment 5

Method

Design

Participants

Stimuli and apparatus

Procedure

Results

Imputed velocity model

Summary

Experiment 6
Method ........................................................................................................ 80
Design ........................................................................................................ 80
Participants ................................................................................................. 80
Stimuli and apparatus .................................................................................. 81
Procedure ..................................................................................................... 81
Results .......................................................................................................... 81
Imputed velocity model .............................................................................. 85
Comparison of results across standard velocities ...................................... 85
Summary ....................................................................................................... 86
General Discussion ....................................................................................... 86

PART II

CHAPTER 4: ATTENTIONAL TUNING TO FREQUENCY AND TIME ............... 93
Attentional Tuning to Frequency ................................................................. 94
A brief overview of function of the peripheral auditory system ............... 94
The probe-signal method ........................................................................... 96
Attentional Tuning to Time ........................................................................ 101
Attentional Tuning to Frequency and Time ............................................. 107

CHAPTER 5: ASSESSING EFFECTS OF EXPECTATION ON PERCEPTUAL
SENSITIVITY ............................................................................................... 114
Experiment 7 ............................................................................................... 114
Overview ...................................................................................................... 114
Method ........................................................................................................ 115
Design ......................................................................................................... 115
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants</td>
<td>115</td>
</tr>
<tr>
<td>Stimuli and apparatus</td>
<td>115</td>
</tr>
<tr>
<td>Procedure</td>
<td>116</td>
</tr>
<tr>
<td>Data analysis</td>
<td>116</td>
</tr>
<tr>
<td>Results and Discussion</td>
<td>118</td>
</tr>
<tr>
<td>Simulation 1</td>
<td>120</td>
</tr>
<tr>
<td>Results and Discussion</td>
<td>122</td>
</tr>
<tr>
<td>Simulation 2</td>
<td>126</td>
</tr>
<tr>
<td>Results and Discussion</td>
<td>128</td>
</tr>
<tr>
<td>Summary</td>
<td>129</td>
</tr>
<tr>
<td>CHAPTER 6: FREQUENCY-TIME INTERDEPENDENCE IN ATTENTION</td>
<td>131</td>
</tr>
<tr>
<td>Experiment 8</td>
<td>131</td>
</tr>
<tr>
<td>Method</td>
<td>131</td>
</tr>
<tr>
<td>Design</td>
<td>131</td>
</tr>
<tr>
<td>Participants</td>
<td>131</td>
</tr>
<tr>
<td>Stimuli and apparatus</td>
<td>132</td>
</tr>
<tr>
<td>Procedure</td>
<td>132</td>
</tr>
<tr>
<td>Data analysis</td>
<td>133</td>
</tr>
<tr>
<td>Results and Discussion</td>
<td>133</td>
</tr>
<tr>
<td>Experiment 9</td>
<td>135</td>
</tr>
<tr>
<td>Overview</td>
<td>135</td>
</tr>
<tr>
<td>Method</td>
<td>138</td>
</tr>
<tr>
<td>Design</td>
<td>138</td>
</tr>
</tbody>
</table>
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Points of Subjective Equality (PSEs) for Experiments 1, 2, and 3</td>
<td>44</td>
</tr>
<tr>
<td>2</td>
<td>Just Noticeable Differences (JNDs) for Experiments 1, 2, and 3</td>
<td>44</td>
</tr>
<tr>
<td>3</td>
<td>Points of Subjective Equality (PSEs) for Experiments 4, 5, and 6</td>
<td>74</td>
</tr>
<tr>
<td>4</td>
<td>Just Noticeable Differences (JNDs) for Experiments 4, 5, and 6</td>
<td>75</td>
</tr>
<tr>
<td>5</td>
<td>Simulation 2. Maximum tolerable $n_A:n_B$ given $n_A$ and true $d'$</td>
<td>128</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Schematic of sounds moving through a two-dimensional pitch-time plane</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>Canonical stimulus and task for the kappa effect</td>
<td>27</td>
</tr>
<tr>
<td>3</td>
<td>Canonical stimulus and task for the tau effect</td>
<td>29</td>
</tr>
<tr>
<td>4</td>
<td>Results from Henry and McAuley (in press)</td>
<td>32</td>
</tr>
<tr>
<td>5</td>
<td>Stimulus diagram for Experiment 1</td>
<td>38</td>
</tr>
<tr>
<td>6</td>
<td>Experiment 1 results</td>
<td>41</td>
</tr>
<tr>
<td>7</td>
<td>Stimulus diagram for Experiment 2</td>
<td>48</td>
</tr>
<tr>
<td>8</td>
<td>Experiment 2 results</td>
<td>49</td>
</tr>
<tr>
<td>9</td>
<td>Stimulus diagram for Experiment 3</td>
<td>54</td>
</tr>
<tr>
<td>10</td>
<td>Experiment 3 results</td>
<td>56</td>
</tr>
<tr>
<td>11</td>
<td>Experiment 1 replication results</td>
<td>64</td>
</tr>
<tr>
<td>12</td>
<td>Stimulus diagram for Experiment 4</td>
<td>69</td>
</tr>
<tr>
<td>13</td>
<td>Experiment 4 results</td>
<td>72</td>
</tr>
<tr>
<td>14</td>
<td>Stimulus diagram for Experiment 5</td>
<td>78</td>
</tr>
<tr>
<td>15</td>
<td>Experiment 5 results</td>
<td>79</td>
</tr>
<tr>
<td>16</td>
<td>Stimulus diagram for Experiment 6</td>
<td>82</td>
</tr>
<tr>
<td>17</td>
<td>Experiment 6 results</td>
<td>83</td>
</tr>
<tr>
<td>18</td>
<td>Experiment 4 replication results</td>
<td>89</td>
</tr>
<tr>
<td>19</td>
<td>(A) Results from a probe-signal experiment; (B) An expectancy profile</td>
<td>98</td>
</tr>
<tr>
<td>20</td>
<td>(A) An isolated-interval duration judgment task; (B) An embedded-interval duration judgment task</td>
<td>102</td>
</tr>
</tbody>
</table>
Results from Crum and Hafter (2008) ................................................................. 109
Stimulus and task diagram for Experiments 7 and 8 ........................................... 117
Experiment 7 results ........................................................................................ 119
Simulation 1 results, MacMillan and Creelman (2005) correction...................... 123
Simulation 1 results, Hautus (1995) correct ...................................................... 124
Simulation 2 results.......................................................................................... 130
Experiment 8 results ...................................................................................... 134
Stimulus and task diagram for Experiment 9
(A) Ascending glide cues .............................................................................. 136
(B) Descending glide cues ........................................................................... 137
Experiment 9 results, perceptual sensitivity .................................................... 142
Experiment 9 results, weighted measure (WM) ............................................. 144
Stimulus and task diagram for Experiment 10
(A) Ascending sequence cues ................................................................. 147
(B) Descending sequence cues ................................................................ 148
Experiment 10 results, perceptual sensitivity ................................................ 152
Experiment 10 results, weighted measure (WM) ........................................... 155
(A) Stimulus from Tougas and Bregman (1985, 1990); (B) Stimulus from
McPherson et al. (1994)............................................................................. 181
INTRODUCTION

Human listeners must solve a very difficult problem when it comes to making sense of their auditory environment. Namely, they must parse the disorganized acoustic information arriving at the ears into separate auditory objects, or streams, corresponding to for example, a friend’s voice, music from a nearby radio, and background sounds coming from a fan or traffic. Although human listeners succeed at this task without effort, how they accomplish such a feat is not well understood. In this regard, there is a long history of psychological research focused on elaborating the ‘bottom-up’ acoustic cues and ‘top-down’ knowledge-based cues that contribute to a listener’s ability to distinguish different sound sources (Broadbent, 1958; Deutsch & Deutsch, 1963; Treisman, 1960; Treisman & Geffen, 1967). At issue is how a listener decides which parts of the incoming sensory stimulation correspond to the same auditory object or event (Bregman, 1990, p. 3). This is the problem of auditory scene analysis.

Perhaps the most widely-cited theory of auditory scene analysis is due to Bregman (1990). Bregman’s theory distinguishes between ‘primitive’, pre-attentive stream segregation principles that are roughly analogous to Gestalt grouping rules from vision, and ‘schema-driven’ stream segregation, which operates based on effortful attentional processes and knowledge of regularities picked up through experience with the auditory world. With respect to primitive stream segregation, sounds tend to be grouped based on, for example, pitch proximity and timbre similarity. That is, listeners are more likely to attribute auditory serial patterns with like pitches or like timbres, as opposed to patterns with disparate pitches or timbres, to a single source. In contrast, Bregman assumes that pitch and time regularities of a sequence operate at the schema-driven level.

Throughout this dissertation, I will refer to pitch-time regularities or pitch-time structure.
Generally, time regularity refers to the tendency of information-carrying events comprising serial auditory patterns to occur at equal-duration time intervals; strict repetition of a time interval is referred to as isochrony. The most obvious example of temporal regularity is musical rhythm, but speech and environmental sounds can also be rhythmic. Pitch regularity describes the tendency of each time interval to be associated with an approximately equal-sized pitch interval. Thus, if sounds are represented as moving through a two-dimensional pitch-time plane, those with a high degree of pitch-time regularity will approximate a straight line through the space (see Figure 1). Restated, highly regular sounds are assumed to approximate constant velocity, where velocity is understood as the change in pitch per change in unit time of an auditory stimulus.

Figure 1. Two sounds represented in a two-dimensional pitch-time plane. Sound A (bold line) exhibits a high degree of regularity and is thus a straight line through the space. Sound B (dotted line) is irregular and so is represented by a jagged line through the space.
Auditory Motion for Continuous and Discrete Sounds

The aim of this dissertation is to test the general hypothesis that, contrary to Bregman’s view, regular pitch-time trajectories (i.e., constant velocities) facilitate perception of and attention to auditory stimuli. Previous research indicates that auditory stimuli occurring at expected points in time (Barnes & Jones, 2000; Jones, Moynihan, MacKenzie, & Puente, 2002; Large & Jones, 1999; McAuley & Kidd, 1998) or at expected locations in pitch space (Greenberg & Larkin, 1968; Hafter, Sarampalis, & Loui, 2007; Johnson & Hafter, 1980; Schlauch & Hafter, 1991) are better perceived than auditory stimuli occurring at unexpected times or pitches. Here, the focus is on the influence of expectations generated based on joint pitch-time trajectory information on perception of and attention to continuous and discrete auditory events. It is assumed that when expectations are satisfied, listeners are prepared, and thus in a better position to process new information. However, when expectations are violated, perceptual distortions are often observed (Henry, 2007; Henry & McAuley, 2009; Henry, McAuley, & Zaleha, 2009; Jones, Maser, & Kidd, 1978; MacKenzie, 2007).

The dissertation is organized into three parts. Part I consists of Chapters 1–3, which focus on pitch-time interactions in perception. The general question addressed in these chapters is whether perceived duration and pitch-change of a to-be-judged stimulus depend on expectations generated based on constant velocity information. Chapter 1 reviews data and theory related to two contrasting perspectives on pitch-time relations; one view assumes that pitch and time are perceptually independent dimensions, while the second assumes that pitch and time are by nature interdependent. An auditory motion hypothesis is outlined that provides a basis for predicting the nature of pitch-time interactions in perception. In Chapter 2, three experiments are reported that provide a test of the auditory motion hypothesis. In all three experiments, listeners judged either the duration or total pitch change of a comparison tone glide.
relative to a standard tone glide, where the comparison’s velocity varied on a trial to trial basis relative to the fixed standard velocity. In Chapter 3, three experiments are reported which extend the results of the experiments reported in Chapter 2 to discrete tone sequences.

Part II consists of Chapters 4 – 6, which focus on pitch-time interactions in attention. The primary question is whether listeners use the pitch-time trajectory of a constant-velocity cue to tune attention to a future pitch-time location at which a signal is expected to occur. Chapter 4 provides a review of data and theory related to attentional tuning. Specifically, previous research indicates that attention can be tuned to locations in pitch and in time; the limited research supporting the possibility that attention is tuned by pitch-time trajectories is also reviewed. Chapter 5 addresses an important methodological issue that arises in assessments of attentional tuning, where detection of an expected event is pitted against detection of unexpected events; specifically, when expected events are defined as those events which occur most often, differences in estimated sensitivity may be observed between expected and unexpected conditions that do not reflect differences in true sensitivity. One experiment and two simulations are reported. In Chapter 6, three experiments are reported that test the hypothesis that listeners better detect signals that occur at a pitch-time location that is expected based extrapolation of the trajectory of a continuous glide cue (Experiment 10) or a discrete sequence cue (Experiment 11).

Finally, Part III consists of Chapter 7, which provides a summary of conclusions and implications suggested by the results of the current experiments.
PART I
CHAPTER 1: PERSPECTIVES ON THE PERCEPTUAL INTERDEPENDENCE OF PITCH 
AND TIME

A fundamental question in cognitive science concerns the way that human listeners parse 
their auditory environment into separate auditory streams attributable to separate sound sources. 
Initially, the stimulation caused by environmental events is organized only insofar as receptors in 
the inner ear respond selectively to frequency. However, the tonotopic organization of the 
peripheral auditory system is insufficient for the listener to deduce which frequency components 
originated from which sound source. The problem of auditory scene analysis is that of 
understanding what cues listeners use to assign temporally extended patterns of auditory 
stimulation to a single sound source. Previous research indicates that listeners make use of both 
pitch and time information to do so (Bregman & Rudnicky, 1975; Handel, Weaver, & Lawson, 
1983; Jones, Kidd, & Wetzel, 1981); the view taken here is that naturally occurring sounds often 
have built-in pitch and time regularities that afford generation of expectations about the future 
time-course of an auditory event (Jones, 1976, 2004), and auditory streams can and are formed 
on the basis of these regularities. Here, pitch and time regularities refer to ordered relationships 
between pitch intervals and/or time intervals in auditory stimuli that give rise to a predictable 
pitch-time trajectory. One question that arises from this view concerns the nature of the interplay 
between the perceptual dimensions of pitch and time in attention and perception, which has been 
the subject of debate (Boltz, 1989b, 1993; Henry & McAuley, 2009; Henry, et al., 2009; Jones, 
Johnston, & Puente, 2006; Jones & Yee, 1993; Monahan & Carterette, 1985; Palmer & 
Krumhansl, 1987; Peretz & Coltheart, 2003; Peretz & Morais, 1989). It should be noted that 
research on pitch-time relations has been conducted largely in a musical domain; Appendix A 
provides the reader with a glossary of key terms, mostly musical, and descriptions of popular
neuroimaging methodologies in this field.

In general, two theoretical perspectives can be contrasted with respect to the relative independence of pitch and time processing. On the one hand, Peretz and colleagues have argued on the basis on reports of amusia (tone deafness) that is either congenital or acquired as the result of a brain injury that pitch and time are independent perceptual dimensions (Peretz & Coltheart, 2003; Peretz & Kolinsky, 1993). Some neural evidence supports this proposition, suggesting the involvement of a right-lateralized fronto-temporal network in the processing of pitch (Hyde, Peretz, & Zatorre, 2008; Hyde, et al., 2009; Hyde, et al., 2007; Hyde, Zatorre, Griffiths, Lerch, & Peretz, 2006; Loui & Schlaug, 2009; Peretz, Brattico, Jarvenpaa, & Tervaniemi, 2009; Peretz, Brattico, & Tervaniemi, 2005) and a more distributed network underlying temporal processing (Di Pietro, Laganaro, Leemann, & Schnider, 2004; Robin, Tranel, & Damasio, 1990; Stewart, Von Kriegstein, Warren, & Griffiths, 2006).

On the other hand, pitch and time have been argued to be fundamentally integral dimensions, with temporal properties of stimuli partially determining perceived pitch and vice versa (Jones, 1976; Jones & Yee, 1993). Support for this view comes from demonstrations that processing of and recall for music is improved when pitch and time accents are aligned (Boltz, 1989b; Boltz & Jones, 1986; Boltz, 1992), and from work suggesting that pitch and timing are better perceived when a target event is expected with respect to both the to-be-judged and to-be-ignored dimensions. Moreover, some intriguing recent neural evidence is suggestive of shared neural circuitry for pitch and time perception (Alain, Cortese, & Picton, 1998; Griffiths, Johnsrude, Dean, & Green, 1999) and superadditive benefits with respect to attentional processes when pitch and time are jointly predictable (Neuhaus & Knösche, 2008); the latter is suggested by EEG work indicating that an early component associated with attentional processes (i.e.,
N100) is attenuated when musical stimuli are predictable in terms of either pitch or time, and that the effect is interactive – the largest N100 attenuation occurs when both pitch and time are predictable (Neuhaus & Knösche, 2008). In the remainder of this chapter, both perspectives on perceptual interdependence will be outlined in more detail, and behavioral and neuropsychological data supporting both perspectives will be reviewed.

**Perceptual Independence of Pitch and Time**

One view of the relationship between pitch and time in perception proposes strict independence in terms of both processing and neural organization (Peretz & Coltheart, 2003). Behavioral evidence for this view comes largely from studies of congenital and acquired amusia. With respect to brain organization, lesion studies and imaging work support structural separation of pitch versus time processing. I will address behavioral and neural evidence in turn.

*Behavioral evidence for pitch-time independence*

Amusia (tone deafness) refers to lifelong impairment in musical ability that is unrelated to hearing acuity, general cognitive functioning, or exposure to music (Ayotte, Peretz, & Hyde, 2002). Amusia can be either congenital, that is, present from birth, or acquired as the result of brain injury. The key feature of amusia is that the associated musical impairments have been primarily linked to pitch processing (Peretz, et al., 2002; Peretz & Hyde, 2003); reports of rhythm or speech perception impairments in amusic individuals are relatively rare (Ayotte, et al., 2002; Hyde & Peretz, 2004; however, see Dalla Bella & Peretz, 2003; Foxton, Nandy, & Griffiths, 2006; Hutchins, Gosselin, & Peretz, in press; Patel, Foxton, & Griffiths, 2005; Patel, Wong, Foxton, Lochy, & Peretz, 2008).

The hallmark of congenital amusia is impaired ability to perceive fine-grained pitch changes [less than ~1-2 ST (Hyde & Peretz, 2004; Peretz, et al., 2005)]; Foxton, Dean, Gee,
Peretz, and Griffiths (2004) showed that amusics performed significantly worse than control listeners on tasks that required detection of pitch differences between pairs of tones, detection of continuous pitch change in tone glides, and determining the direction of a tone glide’s pitch change. Moreover, in a same-different melody comparison task, amusic listeners were less able than controls to identify pitch contour violations (in transposed and non-transposed melodies) and contour-preserving pitch shifts of single notes (i.e., pitch interval disruptions) in short four-tone sequences.

The consequences of the basic pitch perception impairment are suggested to impact performance on a range of more complex listening tasks (Hyde & Peretz, 2004). For example, Ayotte et al. (2002) showed that amusics were less able than controls to detect out-of-tune notes in either familiar or unfamiliar melodies. Amusic listeners were also shown to be less sensitive to musical dissonance than controls, where musical dissonance refers to the sense of unpleasantness experienced by a listener when two notes are played simultaneously that have a complex ratio relationship (for example, two neighboring keys on a piano). However, for affective ratings of major and minor musical excerpts (happy vs. sad), amusics’ responses showed the same pattern as controls, despite the attenuated magnitude of happy / sad ratings relative to control listeners. Thus the abilities of amusic listeners to extract fine-grained pitch information from melodies is impaired, while the ability to glean emotional information remains largely intact.

Musical memory has also been shown to be impaired in tone deafness, with amusic listeners demonstrating an inability to recognize familiar melodies without the aid of lyrics and an impairment on an incidental learning task using unfamiliar melodies (Ayotte, et al., 2002). However, in the same study, amusic listeners were able to recognize the same familiar melodies as well as controls based on lyrical content, and performed normally on an incidental learning
task using environmental sounds and human voices, providing evidence for the specificity of amusia to musical pitch.

In contrast to pitch processing, temporal processing is proposed to be intact in amusic listeners. Supporting this claim, Hyde & Peretz (2004) showed unimpaired temporal processing in amusics in a task that required detection of a tone that occurred early or late relative to its expected temporal position in an otherwise isochronous sequence. However, it should be noted that there is some evidence that amusics are impaired relative to controls in tasks that probe processing of temporal organization (e.g., Peretz, et al., 2002). For example, Foxton et al. (2006) tested amusic and control listeners on a task that involved detection of a rhythmic perturbation in sequences that were either isochronous or had a simple rhythmic pattern (short-long-short-long). Sequences were either monotone or individual tones took on random pitches. Amusics’ temporal discrimination thresholds were significantly worse than controls in all conditions in which pitch was varying, but amusics performed similarly to controls when sequences were monotonic. Moreover, Ayotte et al. (2002) and Dalla Bella and Peretz (2003) have shown that amusics are impaired at tapping along with music relative to controls. Thus, amusics perform as well as controls on time tasks only when the stimuli do not contain dynamic pitch changes.

Amusics have also been proposed to be unimpaired when judging pitch changes in speech material; this is taken as further evidence for independence of auditory processes such as pitch and time perception, but goes a step further and suggests a distinction between music and speech processing. Ayotte et al. (2002) altered the pitch of sentence-final words in short utterances, and listeners judged whether the spoken sentence corresponded to a question (rising final pitch) or a statement (falling final pitch). In this study, the pitch manipulations to the speech material were much larger (averaging 7 ST and 12 ST for statements and questions, respectively)
than the manipulations to musical material that have been shown to be difficult for amusic listeners (i.e., ≤ 1 ST). However, when linguistic information was removed from the utterances, so that stimuli were non-speech analogues of the speech material, amusics were found to be impaired relative to controls on a task that required a ‘rising’ / ‘falling’ judgment, despite the relatively large pitch manipulations. In contrast, Patel and colleagues (Patel, Foxton, et al., 2005; Patel, et al., 2008) have demonstrated using the same paradigm that tone-deaf listeners are impaired relative to controls on a task that requires discriminating a question from a statement based on a rising versus falling final pitch, even when the pitch changes are quite large (up to 21 ST).

To summarize, the findings reviewed above suggest that congenital amusia is characterized by problems perceiving pitch information in musical stimuli (Hyde & Peretz, 2004; Peretz, et al., 2002), but unimpaired rhythm / time perception, interpretation of musical affect (Hyde & Peretz, 2004), recognition of / memory for environmental sounds (Ayotte, et al., 2002), and perception of speech prosody (Ayotte, et al., 2002). Thus, this work supports the independence of melodic pitch processing from other aspects of auditory perception, including temporal processing. However, it should be noted that some results are mixed, and amusic listeners have been shown, in some cases, to perform relatively poorly on tests of temporal processing and speech intonation perception (Dalla Bella & Peretz, 2003; Foxton, et al., 2006; Patel, Foxton, et al., 2005; Patel, et al., 2008). Next, I will review studies of individuals with acquired musical deficits resulting from localized brain lesions and discuss recent neuroimaging work that provides evidence for the neural specialization of pitch and time processing.

**Neural evidence for pitch-time independence**

**Acquired amusia (lesion studies).** Reports of lesions resulting from brain injuries with
associated musical impairments have been informative in terms of assigning musical functions to specific brain regions. Particularly relevant for the current discussion are reports that demonstrate a dissociation of pitch and time processing. In general, the strongest evidence for dissociations comes from reports of unilateral brain lesions; unilateral damage to the right hemisphere is most often implicated in pitch processing deficits, while unilateral left or bilateral damage is more often associated with temporal processing impairments.

Robin et al. (1990) tested patients with left or right unilateral lesions to tempoparietal areas on pitch and time tasks. Pitch tasks included identification of a deviant pitch in a three-interval forced choice task, and adjusting the pitch of a comparison stimulus to match the pitch of a standard. Time tasks included detection of a short silence in a filled interval (i.e., gap detection) and detection of a relatively short inter-stimulus interval in an otherwise isochronous sequence. Patients with left-hemisphere damage performed worse on the time tasks than the pitch tasks, while patients with right-hemisphere damage performed worse on the pitch tasks than on the time tasks. Consistent with the work of Robin et al. (1990), Zatorre (1988) showed that impairments in complex pitch perception (i.e., perception of the missing fundamental in a complex tone) were associated with lesions to the right temporal lobe that encroached on Heschl’s gyrus. Listeners with left temporal lobe damage or damage to the right leaving Heschl’s gyrus intact performed similarly to controls on the complex pitch perception task.

Several case studies provide converging evidence that pitch processing is dependent on right-hemisphere structures and time processing on left-hemisphere structures. Murayama, Kashiwagi, Kashiwagi, and Mimura (2004) evaluated singing performance of one patient before and after an injury to a large volume in right hemisphere including superior temporal gyrus, premotor and primary motor cortices, frontal cortex, and inferior parietal cortex. The patient
demonstrated impaired pitch interval production but preserved rhythm production; however, pitch contour production was reported to be relatively preserved. Di Pietro (2004) described a musician who, after suffering damage to superior and middle temporal gyri and inferior parietal lobule on the left, was unable to reproduce rhythms presented in the auditory modality, but performed perfectly on tasks involving pitch perception and memory, including singing a familiar song from memory, note singing and note naming from sheet music, and discriminating pitches and timbres. Notably, the patient also performed perfectly on tasks requiring production of a rhythm from a written score and discriminating durations.

Taken together, these studies indicate that unilateral brain damage often leads to selective impairments in either pitch or time processing, leaving intact other faculties. However, it is not clear that pitch and time processing are always cleanly dissociated in lesion studies. In this regard, one undeniable obstacle to drawing inferences on the basis of lesion studies is the inherently uncontrolled nature of the location and extent of damage. Studies of patients following brain damage are valuable because they provide information about what faculties are, in principle, dissociable. However, individual differences such as handedness and extent of musical training have implications for brain organization on an individual level, and thus make more difficult inference drawing at a more global level (Stewart, et al., 2006). One way to avoid these criticisms is to make use of neuroimaging techniques on normal (non-patient) listeners and individuals with congenital amusia. In this way, individual difference factors can be more tightly controlled, thereby increasing the inference-drawing power of these studies. In what follows, I will provide a review of studies representative of this approach, and outline the conclusions they have to offer.

*Neuroimaging studies.* Neurorimagining studies can be roughly divided into two types (see
Appendix A for a glossary of neuroimaging techniques). On the one hand, comparisons of congenitally amusic listeners and control listeners using neuroimaging technologies have allowed for detailed analysis of the structural and functional characteristics that make the brains of amusics different. On the other hand, studies using normal (non-patient) listeners focus on comparing brain activation in conditions in which individuals make judgments about the pitch versus the time aspects of auditory stimuli.

Structural MRI studies of individuals with amusia have failed to reveal any gross structural abnormalities that would distinguish these listeners from non-amusics (e.g., Peretz, et al., 2002). However, increasingly sensitive neuroimaging techniques have more recently begun to reveal subtle differences between amusic and non-amusic listeners in terms of connectivity and cortical thickness in auditory and frontal areas. Hyde and colleagues (2007) examined the brains of individuals with congenital amusia using voxel-based morphometry, and found that individuals with amusia tended to have increased cortical volume (i.e., gray matter) in right inferior frontal cortex and right auditory area BA22 in temporal cortex. Moreover, these same individuals were shown to have decreased connectivity (i.e., white matter) in right inferior frontal areas (Hyde, et al., 2006), and white matter concentration was positively correlated with performance on the Scale subtest of the Montreal Battery of Evaluation of Amusia (MBEA; see Appendix A). Finally, Loui and Schlaug (2009), using diffusion tensor imaging, observed reduced white matter in the arcuate fasciculus, a fiber tract connecting auditory and frontal areas, especially on the right. Arcuate fasciculus volume was negatively correlated with behavioral pitch discrimination thresholds.

Functional differences between amusic and control listeners have been shown using EEG technologies. Peretz et al. (2005) observed abnormal ERP components in amusic listeners in
response to a deviant pitch in an otherwise monotone sequence. In particular, N200 and P300 components differentiated amusic from non-amusic listeners, and localization techniques implicated abnormalities in dynamics of the same right frontal-temporal network associated with impaired pitch perception, consistent with the reported MRI studies.

Functional imaging studies of non-amusic listeners provide converging evidence for neural specificity of pitch and time processing. Zatorre, Evans, and Meyer (1994) compared brain activation during the presentation of novel melodies to a matched noise presentation condition using PET. They observed activation in right superior temporal gyrus anterior to primary auditory cortex for the melody presentation condition. In an fMRI study of pitch discrimination, Hyde, Peretz, and Zatorre (2008) examined activation in left and right primary auditory cortices of non-amusic listeners. Listeners heard 21-tone sequences with identical pitch contour, but pitch distances between individual tones were varied parametrically (0, 6.25, 12.5, 25, 50, 100, or 200 cents$^1$). Activation in right planum temporale (located in primary auditory cortex) was positively correlated with the magnitude of pitch distance between tones in the stimulus sequences. However, activation in the analogous area on the left did not change as a function of pitch distance, with the exception of an increase in activation for the largest pitch distance condition (200 cents) relative to the other conditions. Similarly, Zatorre and Belin (2001) repeatedly presented tone pairs to listeners, where the timing and pitch distance between events was varied parametrically in separate blocks. Timing variations involved speeding up the rate of presentation, thus requiring finer temporal resolution at faster rates. The pitch distance between tones was fixed at one octave. Pitch variations involved decreasing the pitch distance between tones with presentation rate held constant at 667 ms, thus requiring finer pitch resolution. Activation in the left temporal cortex was correlated with presentation rate, while
activation in the right temporal cortex was correlated with pitch distance.

To summarize, neuroimaging work on amusic and non-amusic listeners supports conclusions drawn from observations of brain-damaged patients, and indicates that processing of pitch and time components of auditory stimuli is functionally separable (Hyde, et al., 2006; Loui & Schlaug, 2009; Peretz, et al., 2005; Zatorre, et al., 1994). Accumulating evidence is suggestive of a trade-off in the pitch versus time resolution capable by the right and left hemispheres. Pitch and time processing are often lateralized to the right and left brain hemispheres, respectively (Hyde, et al., 2008; Zatorre & Belin, 2001). Thus damage or incomplete development of the right-lateralized pitch-processing network is suggested to underlie the pattern of impaired pitch perception and intact time perception characterizing amusia (Hyde, et al., 2007; Hyde, et al., 2006; Peretz, et al., 2005).

Summary

Support for independence of pitch and time in perception comes from studies of congenital and acquired amusia, and from neuroimaging work on both amusic and non-amusic listeners. Dissociations between musical faculties have been demonstrated in amusia; for example, it is common to observe profoundly impaired pitch processing but relatively preserved time processing. The consensus from behavioral and neuroimaging work on individuals with amusia is that a right fronto-temporal network underlies processing of pitch in music, and amusia can result from damage or incomplete development of this network. Next, the data and theory associated with a pitch-time interdependence perspective will be reviewed.

Perceptual Interdependence of Pitch and Time

An alternative view of auditory processing proposes that pitch and time are fundamentally integrated dimensions; that is, time properties of auditory stimuli partially
determine perceived pitch and vice versa. Evidence for this perspective has come largely from behavioral studies that show better perception of and memory for musical stimuli when pitch and time accents coincide (Boltz, 1989a, 1989b, 1993; Boltz & Jones, 1986; Jones & Boltz, 1989) and increased sensitivity to time or pitch deviations when the to-be-judged event is predictable along the irrelevant dimension (Henry & McAuley, 2009; Henry, et al., 2009; Jones, Boltz, & Kidd, 1982; Jones, et al., 2002). Recent neuroimaging work supports shared neural substrates underlying pitch and time processing (Alain, et al., 1998; Griffiths, et al., 1999) and interactivity of early attentional processes when pitch and time are jointly processed (Neuhaus & Knösche, 2008). I will address in turn the behavioral and neuroimaging evidence for perceptual interdependence view. Then, an auditory motion hypothesis (Henry & McAuley, 2009; Henry, et al., 2009; Jones, 1976; Jones & Yee, 1993; MacKenzie, 2007) will be presented, which suggests that listeners make use of velocity information conveyed by stimuli that move or imply motion in pitch space.

Behavioral evidence for pitch-time interdependence

Support for pitch-time interdependence in perception comes in part from work on joint accent structure; joint accent structure refers to the relation between accents in music marked by pitch or timing changes (see Appendix A); for example, a pitch accent can be signaled by a reversal in pitch direction (a contour change) or the presence of a relatively large pitch interval, and a time accent can be marked by a lengthened note or a period of silence (a rest). In this regard, pitch and time accents can occur either coincidently or asynchronously. Joint accent structure is considered concordant when pitch and time accents occur coincidently, and discordant when they occur asynchronously. Work in this area shows that musical patterns are better perceived, remembered, and reproduced when pitch and time accents in musical stimuli
are concordant (Boltz, 1998; Boltz & Jones, 1986; Jones, 1987; Jones & Ralston, 1991; Jones & Boltz, 1989; Jones, et al., 1982). For example, Boltz (1998) asked listeners to attend to time, pitch, or neither in novel melodies that were either concordant or discordant in terms of accent structure, with the expectation that they would complete a memory test regarding the to-be-attended dimension (listeners not instructed to attend to either dimension were not told to expect a memory test). All listeners were then asked to reproduce the total duration of the melodies; the duration test was a surprise for listeners instructed to attend to pitch or given no instructions. Overall, the durations of concordant melodies were most accurately reproduced. Moreover, for concordant melodies, reproduction accuracy was equally good for conditions that required attending to time and attending to pitch, even though listeners attending to pitch were not expecting a memory test for duration. On the other hand, for discordant melodies, duration estimates were best for the condition that involved attending to time, intermediate for the no-attending condition, and worst when participants attended to nontemporal (pitch) information.

In a similar set of studies, Jones and colleagues (Jones & Ralston, 1991; Jones, Summerell, & Marshburn, 1987; Kidd, Boltz, & Jones, 1984) familiarized listeners with a set of melodies (learning phase), which were later tested for recognition (test phase). During the learning phase, each melody was paired with only one rhythm. During the test phase, melodies were paired with either the same or different rhythm as during the learning phase, in the latter case disrupting joint accent structure of the original melodies. Listeners were explicitly instructed to ignore rhythm information and focus only on pitch, indicating whether or not they recognized the melody from the learning phase. Recognition performance was impaired when melodies were paired with different rhythms, and in particular listeners were fooled most often when decoy melodies that preserved the original pitch contour were paired with new rhythms.
Moreover, listeners better identified familiar melodies when they were paired with similar
relative to dissimilar rhythms, and joint accent structure was thus more similar to the original
melody. The results of this set of studies provides evidence that listeners are sensitive to the
pitch-time ‘shape’ of melodies, and are unable to attend to either dimension alone (Jones, 1987).

Converging results have been obtained for tasks involving synchronizing finger taps with
music (Jones & Pfordresher, 1997), detecting pitch deviants embedded in novel melodies (Jones,
et al., 1982), and notating music from memory (Deutsch, 1980). Listeners perform better when
joint accent structure is concordant relative to when joint accent structure is discordant. Taken
together, these results indicate that when pitch and time information unfold together in a
meaningful way, the dimensions are processed jointly, whereas when pitch and time information
are not meaningfully related, listeners behave as if the dimensions were independent.

Evidence for pitch-time interactivity comes also from studies varying temporal accent
structure and tonal stability. First, Boltz (1989c) asked listeners to judge the duration of a
comparison melody relative to a standard, where temporal accent structure and the tonal stability
of the final tone varied between the standard and comparison melodies (Experiment 2). When
standard and comparison melodies shared the same accent structure, an effect of tonal stability
was observed such that melody durations were underestimated for a tonally less stable note.
When temporal accent structure differed between standard and comparison melodies, distortions
in perceived duration due to tonal stability were enhanced for unexpected temporal endings (i.e.,
early, late). Second, in a similar experiment (Boltz, 1989a), listeners rated the completeness of
musical phrases, which ended early, on time, or late with respect to the expected ending based on
temporal accent structure, and ended on a tonally stable or unstable note based on the established
tonality. Overall, melodies ending on a stable note were judged to be more complete than
melodies ending on an unstable note. Moreover, on-time endings were judged to be more complete than temporally unexpected endings. Finally, violations of both tonal and temporal expectancies led to the lowest completeness ratings.

Lebrun-Guillaud and Tillman (2005, 2007) have also shown interactions between temporal and tonal features of auditory stimuli. In one study (Lebrun-Guillaud & Tillman, 2007), isochronous three-chord contexts were followed by monotone three-tone sequences, where the middle tone occurred at an unexpected time on 50% of trials. Listeners indicated whether or not they detected a temporal deviation of the middle tone. The pitch relationship of the three-tone sequence to the chord context was varied, so that the tone sequence was either the most stable tonic tone, or was tonally less stable (dominant, mediant, or leading tone). Overall, listeners better detected temporal asynchronies in the tone sequence when the pitch corresponded to the most stable tonic relative to the less stable tonal functions. Differences in performance were observed in terms of both sensitivity and response bias, suggesting that the role of tonal function affected true sensitivity to detect a time change rather than just the willingness to respond as such.

In a second study by Lebrun-Guillaud and Tillman (2005), listeners made same-different judgments about pairs of chord sequences, where the last chord was either tonally stable or unstable and ended early, on time, or late. The pitch dimension (stable vs. unstable), time dimension (early, on time, or late), neither, or both dimensions could differ between the standard and comparison chord sequence. Listeners were asked to make same-different judgments in response to only one dimension, ignoring the other. Regardless of whether the to-be-judged dimension was pitch or time, listeners were slowest to respond to incongruent pairs (i.e., to-be-judged dimension is same and to-be-ignored is different or vice versa).
Tillmann and Bharucha (2002) asked listeners to perform a speeded classification of a chord to indicate whether all individual tones making up that chord had simultaneous onsets or whether one element started early. Target chords were preceded by a chord context such that the target was either harmonically related or unrelated to the context. An influence of harmonic relatedness was observed for asynchrony judgments such that listeners were faster and more accurate to detect a temporal asynchrony when the target chord was harmonically unrelated to the context. In sum, the studies of Tillmann et al. are consistent with perceptual interdependence, demonstrating interactions between temporal and tonal stimulus features.

Finally, Schellenberg, Krysciak, and Campbell (2000) examined the possibility that pitch and rhythm contribute jointly to the emotional quality of a musical piece. Stimuli for the study were melodies judged to convey only one emotion (happy, sad, scary); two pieces per emotion were selected. Three versions of each melody were then created in addition to the original melodies. A pitch-only condition combined melodic information with an isochronous rhythm, a rhythm-only condition combined rhythmic information with a monotonic melody, and a baseline condition paired a monotonic melody with an isochronous rhythm. Listeners judged how much each version conveyed the emotion of the original selection. Overall, pitch was a stronger contributor to perceived emotion than rhythm. However, pitch and rhythm were seen to interact, but the relationship was not the same for all pieces. In particular, pitch and rhythm jointly contributed to ratings for happy and scary pieces, but not sad pieces.

To summarize, behavioral evidence for pitch-time interdependence suggests that listeners jointly encode pitch and time information in music; thus they better perceive, remember, and reproduce musical pieces that have concordant relative to discordant joint accent structures (Boltz, 1998; Boltz & Jones, 1986; Jones, 1987; Jones & Boltz, 1989; Jones, et al.,
1982), and have trouble recognizing familiar pitch contours when they are presented paired with novel rhythms (Jones & Ralston, 1991; Jones, et al., 1987). Moreover, judgments about tone timing are affected by tonal expectations, and judgments of tonality are affected by temporal expectations (Lebrun-Guillaud & Tillman, 2007; Lebrun-Guillaud & Tillmann, 2005; Tillmann & Bharucha, 2002). Finally, the emotional character of musical pieces has been shown to depend jointly on pitch and time properties (Schellenberg, et al., 2000). One question that often emerges in studies of pitch-time interdependence is whether the statistical interaction between pitch and time variables reached significance, that is whether joint effects were non-additive. Notably, Ellis and Jones (2009) have shown that the presence of a statistical interaction depends on the specific analyses and dependent measures employed; thus the presence or absence of a statistical interaction does not imply the presence or absence of a perceptual interaction. In what follows, I will review the neural evidence for pitch-time interdependence.

**Neural evidence for pitch-time interdependence**

Some neuroimaging work has suggested that pitch and time are processed in an integrated way. However, on the whole the possibility of pitch-time interdependence has received little attention in the context of neuroimaging relative to questions about separability of the dimensions. Particularly relevant in the current context, Neuhaus and Knösche (2008) studied ERPs corresponding to attention to auditory sequences with varying degrees of pitch and time structure. Listeners made same-different judgments about melody pairs, where different pairs were mismatched in terms of the final sequence tone. Melodies were presented in their original form, with their original temporal structure paired with randomized pitches, with their original pitch structure paired with randomized durations, or with both dimensions randomized independently. Overall, ERP magnitude was largest for the melodies with random pitch and time
patterns, and smallest when both dimensions were structured (i.e., original melodies), reflecting more effortful processing of musical sequences with reduced predictability. ERP magnitude was intermediate for the two conditions in which one dimension was randomized. Critically, Neuhaus and Knösche observed interactions between pitch and time structure in the magnitude of N100 and P200 ERP components, both of which are notably associated with early perceptual processes.

Using EEG, Alain, Cortese, and Picton (1998) studied mismatch negativity (MMN) in response to pitch and time deviants in repeating tone sequences. MMNs are brain responses to events that violate listener expectations generated based on regularities in a stimulus sequence. Alain et al. (1998) hypothesized that separate representations for pitch and time expectations should be reflected in MMN differences when expectations were violated on either dimension. Instead, MMN topography did not differ for time versus pitch deviants, and implicated involvement of a right-lateralized temporal-frontal network in detection of both deviant types. More generally the results are consistent with listeners encoding pitch and time properties of auditory sequences in a unified way, such that deviants on either dimension elicit similar patterns of activity.

Finally, Griffiths et al. (1999) used fMRI in tandem with a behavioral paradigm in which listeners completed a same-different melody comparison task, where different melodies could differ in terms of their time structure, pitch structure, or both. Listeners were instructed to make their responses based on only one dimension at a time. Both tasks activated a number of task-relevant areas including superior temporal gyrus and planum temporale, cerebellum, inferior frontal cortex, anterior cingulate cortex, and dorsolateral prefrontal cortex. Comparing the time judgment condition to the pitch judgment condition revealed no areas of significant activation,
leading the authors to conclude that a similar network of brain regions is involved in pitch and
time processing. However, it should be noted that the failure to observe areas of activation
resulting from subtraction of conditions does not necessarily mean that such differences do not
exist.

To summarize, the neuroimaging work supporting pitch-time interdependence is
somewhat sparse; however, in general neuroimaging paradigms have not frequently been
combined with paradigms in which pitch-time interdependence is typically observed. However,
the studies reviewed above indicate that early ERP components are attenuated for comparison of
melodies that are highly structured on both the time and pitch dimensions relative to melodies
that are structured along only one or neither dimension (Neuhaus & Knösche, 2008). Moreover,
detection of pitch and time deviants in repetitive tone sequences elicited similar MMN
topographies (Alain, et al., 1998). Finally, an fMRI study failed to reveal differences in areas
involved in pitch and time processing (Griffiths, et al., 1999).

*Auditory motion hypothesis*

The focus of Part I of this dissertation is pitch-time independence in perception. Thus far,
I have reviewed empirical evidence for two opposing views regarding perceptual
interdependence. In the current section, I will outline a theoretical perspective that provides a
basis for the experiments presented in Part I, namely the auditory motion hypothesis (Henry &
proposes that listeners make use of velocity information in auditory stimuli varying in pitch over
time, where velocity, $v$, is given by $\Delta p / \Delta t$; $\Delta p$ refers to the pitch distance traveled and $\Delta t$ refers to
the duration of the movement (or time traveled). This type of motion will be referred to as
motion in pitch space to distinguish it from the motion of objects in physical space. Motion in
pitch space conveys important information to listeners in both speech and musical contexts (Dilley & McAuley, 2008; Frick, 1985; Jones, 1976; Jones & Yee, 1993; Krumhansl, 1991; Werner & Keller, 1994). For example, in speech communication, patterns of rises and falls in pitch (an element of speech prosody) help determine word boundaries, distinguish between questions and statements, disambiguate semantic content, place emphasis, and convey information about the emotional state of the speaker. In music, patterns of pitch change over time convey information about tonality and contribute to perceived accent structure. More broadly, in the general context of an auditory scene, motion in pitch space also enables listeners to predict the future time-course of a stimulus, that is, ‘what’ will happen ‘when’.

An assumption of the auditory motion hypothesis is that pitch represents an auditory analogue to location in visual physical space (Kubovy, 1981; Kubovy & Van Valkenburg, 2001). Thus, it is worthwhile to briefly review the evidence supporting this link. First, an analogy can be made at the neural level between the topographic organization of the response of the retina to light and the tonotopic organization of the response of the basilar membrane to sound. Notably, retinotopic and tonotopic organizations are preserved in visual and auditory cortical areas, respectively (Lund, 1988; Merzenich, Colwell, & Anderson, 1982). Second, parallels can be made between pitch space and physical space in attentional cueing paradigms. Directing visual attention to a location in physical space improves reaction time to detect a visual target at the cued location (Posner, Snyder, & Davidson, 1980; Posner & Petersen, 1990); similarly, directing auditory attention to a particular pitch or pitch range improves detectability of an auditory signal presented at or near the cued pitch location (Hafter, Schlauch, & Tang, 1993; Howard, O'Toole, Parasuraman, & Bennett, 1984; Howard, O'Toole, & Rice, 1986). Third, in stimulus-response compatibility (SRC) pitch-discrimination tasks, listeners are faster and more accurate when the
mapping between relative pitch height and the corresponding response is congruent (high pitch associated with high button) compared to when the mapping is incongruent (Douglas & Bilkey, 2007; Keller & Koch, 2006; Lidji, Kolinsky, Lochy, Morais, & Roosevelt, 2007; Melara & O'brien, 1987; Mudd, 1963; Rusconi, Kwan, Giordano, Umilta, & Butterworth, 2006). Finally, two spatiotemporal illusions – the tau and kappa effects – demonstrated for physical space in the visual (Bill & Teft, 1969; Cohen, Hansel, & Sylvester, 1953; Jones & Huang, 1982) and auditory modalities (Grondin & Plourde, 2007; Sarrazin, Giraudo, & Pittenger, 2007) have also been reported for motion of sounds in pitch space (Cohen, Hansel, & Sylvester, 1954; Crowder & Neath, 1994; Henry & McAuley, 2009; Henry, et al., 2009). These illusions are explored in more detail below. In sum, there is evidence that pitch embodies properties of a spatial dimension, supporting the general hypothesis that listeners’ perception of motion in pitch space shares similarities with perceived motion in physical space.

The auditory motion hypothesis proposes that listeners pick up on velocity over the course of an auditory stimulus moving in pitch space, and use velocity information to generate expectations about the pitch and timing of future events. When these expectations are violated, perceptual distortions result that reflect the degree of interdependence between perceived pitch and time; larger perceptual distortions reflect increasing degrees of pitch-time interdependence. Providing support for the auditory motion hypothesis, Henry and McAuley (2009) presented listeners with three-tone sequences marking out two time intervals (t1 and t2) and two pitch intervals (p1 and p2), with the pitch and timing of the initial and final tones fixed, thus implying a constant velocity (see Figure 2). Pitch and timing of the middle, target tone were varied independently. Listeners judged the relative timing of the target tone, indicating whether t1 < t2 or t1 > t2, ignoring target pitch. Henry and McAuley observed an auditory kappa effect; that is,
perceived time was distorted such that the target tone was heard as occurring early (relative to the temporal mid-point of the sequence, t1 < t2) when the corresponding pitch distance was relatively short (p1 < p2), and the target tone was heard as occurring late (t1 > t2) when the corresponding pitch distance was relatively long (p1 > p2). That is, the perceived duration of t1 was distorted in the direction consistent with listeners generating expectations about target
timing based on assumed constant velocity across the three-tone sequence.

The distortions in perceived time were well accounted for by an imputed velocity model that assumed that the perceived timing of the target tone was an additive combination of the actual timing of the target tone and the expected timing based on constant velocity. The imputed velocity model took the form

\[
t' = wt + (1 - w)t_E
\]

where \(t'\) corresponds to the perceived duration of the first sequence time interval, \(t_1\), and \(t\) corresponds to the actual duration of \(t_1\). The value \(t_E\) is the expected duration of \(t_1\) based on target tone pitch and assumed constant velocity, \(t_E = p/V\). This follows from the equation for velocity, \(V = p/t\). The parameter \(w\) determines the relative contribution of actual time and expected time to the perceived time of the target tone. When \(w = 1\), perceived time is based only on actual time, and no perceptual distortion occurs. However, when \(w < 1\), perceived time is a weighted combination of actual and expected time; thus, perceived time is distorted in the direction consistent with an assumption of constant velocity. The magnitude of the perceptual distortion increases with decreasing values of \(w\).

Henry et al. (2009) also examined the complementary tau effect in a task where listeners heard similar three-tone sequences and judged the perceived pitch of the target tone, indicating whether \(p_1 < p_2\) or \(p_1 > p_2\), ignoring target timing (see Figure 3). Henry and colleagues found that perceived pitch was distorted such that target pitch was heard as relatively low (with respect to the pitch mid-point of the sequence, \(p_1 < p_2\)) when the corresponding duration was relatively short (\(t_1 < t_2\)), and target pitch was heard as relatively high (\(p_1 > p_2\)) when the corresponding duration was relatively long (\(t_1 > t_2\)). That is, perceived target pitch was distorted in the direction consistent with listeners generating expectations about target pitch based on assumed constant constant
velocity across the three-tone sequence.

![Diagram](image_url)

**Figure 3.** Canonical stimulus and task for the auditory tau effect. A three-tone sequence marks out two pitch intervals (p1,p2) and two time intervals (t1,t2). The pitch and timing of the outside two tones are fixed, while the pitch and timing of the middle, target tone vary from trial to trial. Listeners indicate whether p1 < p2 or p1 > p2, ignoring the relative timing of the target tone.

For the tau effect, the imputed velocity model takes the form

\[ p' = wp + (1 - w)p_E \]  

(2)

where \( p' \) corresponds to perceived target pitch relative to the initial sequence tone, p1, and \( p \) corresponds to the actual target pitch. The value \( p_E \) is the expected relative target pitch based on target tone timing and assumed constant velocity, \( p_E = tV \). Similar to the model of the kappa
effect, $w$ determines the relative contribution of actual pitch and expected pitch to the perceived pitch of the target tone. When $w = 1$, perceived pitch is based only on actual pitch, and no distortion occurs. When $w < 1$, perceived pitch is a weighted combination of actual pitch and expected pitch based on constant velocity, and perceived pitch is distorted. The magnitude of the perceptual distortion increases with decreasing $w$ values.

For both the kappa and tau effects, distortions along the to-be-judged dimension were in the direction that would be expected given the value of the to-be-ignored dimension, assuming constant velocity. This is consistent with listeners making use of velocity in discrete tone sequences to generate expectations about pitch and time, and thus is consistent with the auditory motion hypothesis. A further prediction made by the auditory motion hypothesis for discrete sequences is that listeners are more likely to make use of velocity information in faster auditory stimuli. Specifically, the auditory motion hypothesis proposes the existence of a limited range of velocities within which listeners make use of motion information in discrete tone sequences; within this range, listeners are more likely to make use of velocity information in faster stimuli (Jones, 1976; Jones & Yee, 1993). Confirming this prediction, studies by Henry and McAuley (2009; Henry, et al., 2009) demonstrated that the relative contribution of velocity to perceived time or pitch increased with increasing velocity.

The first set of experiments making up this dissertation (Experiments 1 – 3) provides a further test of the auditory motion hypothesis. The aim was to extend the auditory motion hypothesis to sounds moving continuously in pitch space. This was deemed a particularly important goal because Henry and McAuley (in press) have previously shown that listeners seem to rely on different information when estimating velocity of continuous versus discrete auditory stimuli. In their study, listeners provided velocity estimates of continuous tone glides or discrete
tone sequences using a scale ranging from 0 (not changing at all) to 100 (changing very quickly). Importantly, velocity was conveyed in two ways; for a pitch-change condition, velocity conditions were created by holding the total duration of the stimulus constant and varying total pitch-change, while pitch-change was held constant and duration was varied for a time-change condition. The results are reproduced in Figure 4 for continuous (Panel A) and discrete (Panel B) auditory stimuli. For continuous tone glides, pitch-change and time-change information contributed approximately equally to estimates of perceived velocity. However, for discrete tone sequences, only time-change information contributed to velocity estimates, with little to no contribution of pitch-change information. These results highlight potentially important differences in velocity perception for continuous versus discrete stimuli moving in pitch space.

The remainder of Part I provides 1) an extension of the auditory motion hypothesis to sounds moving continuously in pitch space and 2) a direct comparison of continuous and discrete auditory stimuli. In Experiments 1 – 6, listeners completed a task that involved judging either the duration or pitch-change of a variable-velocity comparison stimulus relative to a constant-velocity standard stimulus. In Experiments 1 – 3, stimuli were continuous tone glides; in Experiments 4 – 6 stimuli were discrete tone sequences. It was expected that listeners would generate expectations about the duration and pitch-change of the comparison stimulus using the constant velocity of the standard stimulus, resulting in perceptual distortions. Moreover, it was of interest whether listeners would rely on velocity to the same extent when stimuli moved continuously in pitch space (Exp. 1 – 3) relative to when stimuli only implied motion in pitch space (Exp. 4 – 6).
Figure 4. Results reproduced from Henry and McAuley (in press). Listeners estimated velocity of sounds moving in pitch space on a scale ranging from 0 (not changing at all) to 100 (changing very quickly). (A) For continuous tone glides, both pitch-change and time-change information contributed to velocity estimates. (B) For discrete tone sequences, only-time change information contributed to estimated velocity, with little to no contribution of pitch-change.
CHAPTER 2: APPLICATION OF THE AUDITORY MOTION HYPOTHESIS TO SOUNDS MOVING CONTINUOUSLY IN PITCH SPACE

The experiments reported in this chapter applied the auditory motion hypothesis to sounds moving continuously in pitch. In three experiments, listeners judged either the total duration (time-change task) or total pitch change (pitch-change task) of a variable-velocity comparison tone glide relative to a constant-velocity standard tone glide. For the time-change task, the standard duration was fixed at 1000 ms, and the duration of the comparison varied around 1000 ms (± 5%, 10%, 15%, 20%). For the pitch-change task, the total pitch change of the standard was fixed at 1000 Hz, and the total pitch change of the comparison varied around 1000 Hz (± 5%, 10%, 15%, 20%). Across all experiments, the comparison velocity was variable from trial to trial and took on values of 500 Hz/s, 1000 Hz/s, and 1500 Hz/s. The standard glide velocity was varied across experiments. In Experiment 1, standard velocity was 1000 Hz/s, in Experiment 2, standard velocity was 500 Hz/s, and in Experiment 3, standard velocity was 1500 Hz/s. Fixing the standard velocity was intended to encourage listeners to assume a constant velocity across the standard-comparison glide pair; thus listeners were expected to use the standard velocity to generate predictions about comparison duration or pitch change, leading to perceptual distortions. In what follows, predictions will be derived for all three experiments according to the imputed velocity model (Henry & McAuley, 2009; Henry, et al., 2009).

When judging time, listeners were assumed to use a combination of actual glide duration \( t \) and expected glide duration based on assumed constant velocity and comparison glide pitch change \( t_E = p/V \). When standard and comparison velocities were matched, and pitch change was thus predictive of total duration, actual and expected duration were equivalent \( t = t_E \). Therefore, when the standard and comparison velocities were matched, perceived duration was
predicted to be undistorted. When comparison glide velocity was relatively fast, if listeners relied
in part on expected duration, $t_E$, they were expected to overestimate the duration of the
comparison glide. This is because the fast-velocity comparison glide underwent a larger total
pitch change per unit time than a comparison with the assumed constant velocity; thus expected
duration, $t_E$, would be relatively large, and comparison duration would be overestimated. On the
other hand, when comparison glide velocity was relatively slow, if listeners relied in part on $t_E$, they were expected to underestimate the duration of the comparison glide. This is because the slow-velocity comparison glide underwent a smaller total pitch change per unit time than a comparison with the assumed constant velocity; thus $t_E$ would be relatively small, and comparison duration would be underestimated.

When judging pitch, listeners were assumed to use a combination of the actual pitch change ($p$) and expected pitch change based on assumed constant velocity and total comparison
duration ($p_E = tV$). When standard and comparison velocities were matched, and duration
accurately predicted pitch change, actual and expected pitch change were equivalent ($p = p_E$).
Therefore, when standard and comparison velocities were matched, perceived pitch was
predicted to be undistorted. When comparison glide velocity was relatively fast, if listeners relied
in part on expected pitch change based on constant velocity and duration, they should
underestimate pitch change. This is because the total duration of the fast-velocity comparison
was relatively short compared to the duration of a comparison with assumed constant velocity,
thus expected pitch change, $p_E$, would be relatively small, and comparison pitch change would be
underestimated. Conversely, when comparison glide velocity was relatively slow, if listeners
relied in part on $p_E$, they were expected to overestimate comparison pitch change. This is because
the slow-velocity comparison glide was long relative to a comparison with assumed constant
velocity; thus $p_E$ would be relatively large, and comparison pitch would be overestimated.

Across all experiments, listeners were expected to overestimate comparison duration and underestimate comparison pitch change when comparison velocity was relatively fast, and when comparison velocity was relatively slow, listeners were expected to underestimate comparison duration and overestimate comparison pitch change. In Experiment 1, the velocity of the standard glide was 1000 Hz/s, and the comparison glide velocity varied from trial to trial, taking on one of three values (500 Hz/s, 1000 Hz/s, 1500 Hz/s). The 500 Hz/s comparison velocity was thus relatively slow, and the 1500 Hz/s comparison velocity was relatively fast. Thus, for the 500 Hz/s comparison, listeners were expected to underestimate duration and overestimate pitch change. For the 1500 Hz/s comparison, listeners were expected to overestimate duration and underestimate pitch change. Listeners were expected to accurately judge both duration and pitch change when the velocity of the comparison glide matched that of the standard (i.e., when comparison velocity was 1000 Hz/s).

In Experiment 2, the velocity of the standard glide was fixed at 500 Hz/s, and as in Experiment 1, the comparison glide velocity varied from trial to trial and took on three values (500 Hz/s, 1000 Hz/s, 1500 Hz/s). Thus the standard and comparison velocities matched when the comparison glide velocity was 500 Hz/s, and the 1000 Hz/s and 1500 Hz/s comparison velocities were relatively fast. Perceived duration and pitch change were predicted to be undistorted when the standard and comparison glide velocities matched (i.e., 500 Hz/s). Moreover, for both relatively fast comparison velocity conditions (1000 Hz/s, 1500 Hz/s), listeners were expected to overestimate comparison duration and underestimate comparison pitch change.

In Experiment 3, the velocity of the standard glide was 1500 Hz/s. The standard and
comparison velocities matched when the comparison glide was 1500 Hz/s, and the 500 Hz/s and 1000 Hz/s comparison velocities were relatively slow. Perceived duration and pitch change were predicted to be undistorted when the standard and comparison glide velocities matched (1500 Hz/s). Moreover, for both relatively slow comparison velocity conditions (500 Hz/s, 1000 Hz/s), listeners were expected to underestimate comparison duration and overestimate comparison pitch change.

Experiment 1

Method

Design
The design was a 2 (Task: time-change, pitch-change) x 3 (Comparison Velocity: 500 Hz/s, 1000 Hz/s, 1500 Hz/s) x 8 (Comparison Level: -20%, -15%, -10%, -5%, +5%, +10%, +15%, +20%) mixed-factorial. Task was a between-subjects factor; Comparison Velocity and Comparison Level were within-subjects factors. Participants judged either the duration or total pitch change of a variable-velocity comparison tone glide relative to a constant-velocity standard (see Figure 5).

Participants
Forty individuals (n = 32, female) from a large Midwestern university participated in exchange for course credit in an introductory psychology class. Listeners self-reported normal hearing and had a range of formal musical training (0 – 15 years, $M = 4.1, SD = 4.0$). Listeners were randomly assigned to complete either the time-change task (n = 22) or the pitch-change task (n = 18).

Stimuli and apparatus
Stimuli were pure-tone glides, ramped over the first and last 5 ms to eliminate acoustic
artifacts. Standard glides ascended at a velocity of 1000 Hz/s, and comparison glides ascended at one of three velocities (500 Hz/s, 1000 Hz/s, 1500 Hz/s). For the current experiment, velocity was specified in terms of linear frequency space, i.e., Hz/s. Standard glides were 1000 ms in duration and spanned 1000 Hz from start to finish. For the time-change task (Figure 5A), comparison duration was varied around 1000 ms (± 5%, 10%, 15%, 20%), and for the pitch-change task (Figure 5B), comparison pitch change was varied around 1000 Hz (± 5%, 10%, 15%, 20%). The starting pitch of both standard and comparison glides was randomized from trial to trial, taking on one of three values (476 Hz, 600 Hz, 756 Hz).

Stimuli were generated using MATLAB software (The Mathworks, Inc.). Stimulus generation and response collection were controlled using E-Prime 2.0.8.73 software (Psychology Software Tools, Inc.) running on Dell Optiplex computers. Auditory stimuli were presented at a comfortable listening level (~70 dB) over Sennheiser HD 280 Pro headphones (Old Lyme, CT). Responses were made using a serial response box; ‘shorter’ and ‘more pitch change’ responses were made by pushing a left-side button, while ‘longer’ and ‘less pitch change’ responses were made by pushing a right-side button.

Procedure

Participants heard recorded instructions, then completed an 18-trial training block with corrective feedback. During training, listeners heard only comparison levels of ± 20%, one trial per Comparison Velocity at each starting pitch value. Listeners then completed two experimental blocks with no feedback. On each trial, participants heard a standard-comparison glide pair, and judged whether the comparison was ‘shorter’ or ‘longer’ than the standard (time-change task) or whether the comparison changed ‘more’ or ‘less’ in pitch than the standard (pitch-change task); see Figure 5. The standard glide ascended at a velocity of 1000 Hz /s from one of three starting
(A) Time-change task

Standard Glide: 1000 Hz/s

Comparison Glide: 500 Hz/s, 1000 Hz/s, 1500 Hz/s

1000 ms ± ΔT

1000 Hz/s

1500 Hz/s

1000 Hz/s

500 Hz/s

1000 ms

1000 Hz

(B) Pitch-change task

Standard Glide: 1000 Hz/s

Comparison Glide: 500 Hz/s, 1000 Hz/s, 1500 Hz/s

1000 Hz/s

1500 Hz/s

500 Hz/s

1000 Hz

1000 ms

1000 ms ± ΔP

Figure 5. Stimulus diagram for Experiment 1. Standard glide velocity was 1000 Hz/s. For both the time change and pitch-change tasks, standard glide duration was 1000 ms and standard glide pitch change was 1000 Hz. (A) For the time-change task, ΔT = ± 5%, 10%, 15%, or 20%, and listeners judged whether the comparison was ‘shorter’ or ‘longer’ than the standard. (B) For the pitch-change task, ΔP = ± 5%, 10%, 15%, or 20%, and listeners judged whether the comparison changed ‘more’ or ‘less’ in pitch than the standard.
pitches (476 Hz, 600 Hz, 756 Hz). The comparison glide ascended at one of three velocities (500 Hz/s, 1000 Hz/s, 1500 Hz/s) from one of the same three starting pitches. For the time-change task, the comparison took on one of eight durations, i.e., ± 5%, 10%, 15%, 20%, relative to the standard. For the pitch-change task, the comparison took on one of eight values of pitch change, i.e., ± 5%, 10%, 15%, 20%, relative to the standard.

In each experimental block, listeners provided three responses to each combination of Comparison Velocity and Comparison Level, for a total of 72 trials. Overall, listeners completed a total of 144 trials in two experimental blocks; thus 6 observations were obtained for each Comparison Velocity x Comparison Level combination. Following completion of the experiment, all listeners completed several surveys assessing aspects of musical and linguistic background, understanding, effort, attention to the task, and perceived task difficulty. The experiment lasted approximately 30 minutes.

Data analysis

Proportions of ‘longer’ responses (time-change task) and ‘more pitch change’ judgments (pitch-change task) were determined for each participant for each of the eight values of Comparison Level for each Comparison Velocity, averaged over the two test blocks. Just noticeable differences (JND) and points of subjective equality (PSE) for the resulting psychometric curves were then estimated for each participant using the z-transform method prescribed by Macmillan and Creelman (2005; pp. 219-220). Based on this analysis, JND is a measure of half the distance between the 25th and 75th percentiles of the cumulative response function, and provides a measure of sensitivity to changes in duration or pitch change. PSE is a measure of the median of the cumulative response function, corresponding to the duration level of the comparison glide judged to be ‘longer’ 50% of the time (time-change task) or the pitch-
change level of the comparison judged as ‘more pitch change’ 50% of the time (pitch-change task). Negative PSE indicates a leftward shift in the psychometric function and a tendency to overestimate the duration or pitch change of the comparison glide, whereas positive PSE indicates a rightward shift in the psychometric function and a tendency to underestimate the duration or pitch change of the comparison. Both PSE and JND are reported as percentage (relative) deviations.

Next, proportions of ‘longer’ and ‘more pitch change’ responses were fit with a modified imputed velocity model (Henry & McAuley, 2009; Henry, et al., 2009), where the relative contribution of actual duration, $t$, and expected duration, $t_E$, (or actual pitch change, $p$, and expected pitch change, $p_E$) to perceived comparison duration, $t'$, (or pitch change, $p'$) were determined by the weight parameter, $w$. Values of $w$ quantify the relative contribution of expected duration or pitch change to perceived duration or pitch change, respectively, and thus reflect the degree to which listeners relied on assumed constant velocity when judging time or pitch. When $w = 1$, perceived time / pitch is based only on actual time / pitch, and no perceptual distortion occurs. However, when $w < 1$, perceived time / pitch is a weighted combination of actual and expected time / pitch; thus, perceived time / pitch is distorted in the direction consistent with an assumption of constant velocity. The magnitude of the perceptual distortion increases with decreasing values of $w$. Model fits minimized root mean square error of approximation between actual and expected values of response proportions (see Henry & McAuley (2009) or Henry, et al. (2009) for full details of model fitting procedures). Mean $w$ values are reported for the time-change and pitch-change tasks.

**Results**

Figure 6 shows the proportion of ‘longer’ responses (time-change task, Panel A) and
Figure 6. Experiment 1. (A) Proportions of ‘longer’ responses and (B) proportions of ‘more pitch change’ responses as a function of Comparison Level for the time change and pitch-change tasks, respectively. When standard and comparison velocities were the same (1000 Hz/s), listeners judged duration and pitch change accurately. When the comparison velocity was relatively fast (1500 Hz/s), duration was overestimated and pitch change was underestimated. When the comparison velocity was relatively slow (500 Hz/s), duration was underestimated and pitch change was overestimated.
‘more pitch change’ responses (pitch-change task, Panel B) as a function of Comparison Level for each of the three comparison velocities (500 Hz/s, 1000 Hz/s, 1500 Hz/s). For the time-change task, perceived duration was undistorted when comparison glide velocity was 1000 Hz/s, that is, when the comparison velocity matched the standard velocity. Moreover, as predicted by the auditory motion hypothesis, comparison duration was overestimated when comparison glide velocity was relatively fast (i.e., 1500 Hz/s) and underestimated when comparison glide velocity was relatively slow (500 Hz/s). For the pitch-change task, perceived pitch change was undistorted when the comparison glide and standard glide were presented at the same velocity (i.e., 1000 Hz/s). As predicted by the auditory motion hypothesis, comparison pitch change was underestimated when comparison glide velocity was relatively fast (1500 Hz/s) and overestimated when comparison glide velocity was relatively slow (500 Hz/s).

Supporting this interpretation, a 2 (Task) x 3 (Comparison Velocity) x 8 (Comparison Level) mixed-measures ANOVA on response proportions (‘longer’ and ‘more pitch change’) revealed a main effect of Comparison Level, $F(7,266) = 309.05, MSE = 0.03, p < 0.001, \eta^2_p = 0.89$, a Task x Comparison Level interaction, $F(2,36) = 2.79, MSE = 0.03, p < 0.01, \eta^2_p = 0.07$, a Comparison Velocity x Comparison Level interaction, $F(14,532) = 27.08, MSE = 0.02, p < 0.001, \eta^2_p = 0.42$, and critically, a Task x Comparison Velocity x Comparison Level interaction $F(14,532) = 9.74, MSE = 0.02, p < 0.001, \eta^2_p = 0.20$. The main effect of Comparison Level confirms that proportions of ‘longer’ and ‘more pitch change’ responses increased with comparison duration and pitch change, respectively. That is, listeners responded ‘longer’ more often when the comparison was relatively long, and listeners responded ‘more pitch change’ more often when the comparison pitch change was relatively large. The two-way interactions must be interpreted in the context of the significant three-way interaction (see Figure 6), which is
consistent with the pattern of results predicted by the auditory motion hypothesis. Specifically, comparison duration was overestimated and comparison pitch change was underestimated when comparison glide velocity was relatively fast (1500 Hz/s), while comparison duration was underestimated and comparison pitch change was overestimated when comparison glide velocity was relatively slow (500 Hz/s).

A 2 (Task) x 3 (Comparison Velocity) mixed-measures ANOVA on PSE supported the analysis of response proportions. Specifically, the ANOVA revealed a main effect of Task, $F(1,38) = 5.70$, $MSE = 0.10$, $p < 0.05$, $\eta^2_p = 0.13$, and a Task x Comparison Velocity interaction, $F(2,76) = 25.21$, $MSE = 0.14$, $p < 0.001$, $\eta^2_p = 0.40$. The main effect of Task indicated that, overall, PSEs tended to be slightly larger for the pitch-change task ($PSE = 7.5\% \pm 4.3\%$) than for the time-change task ($PSE = 6.2\% \pm 3.9\%$). However, the main effect of task was qualified by the Task x Comparison Velocity interaction (Table 1), which highlights opposite patterns of perceptual distortions as a function of Comparison Velocity for the time-change task versus the pitch-change task; that is, duration was overestimated when comparison velocity was relatively fast (1500 Hz/s) and underestimated when comparison velocity was relatively slow (500 Hz/s), while the opposite was true for pitch change; pitch change was underestimated when comparison velocity was relatively fast (1500 Hz/s) and overestimated when comparison velocity was relatively slow (500 Hz/s). Single-sample $t$-tests indicated that PSEs for the 1000 Hz/s comparison velocity did not differ from 0 for either the time-change task ($PSE = 0.3\% \pm 1.9\%$, $p = 0.50$) or the pitch-change task ($PSE = 1.4\% \pm 3.6\%$, $p = 0.11$). The main effect of Comparison Velocity did not reach significance, $p = 0.52$. 
A parallel analysis on JNDs revealed only a main effect of Comparison Velocity, $F(2,76) = 11.07, \text{MSE} = 0.03, p < 0.01, \eta^2_p = 0.23$ (see Table 2). Thresholds were lower for the 1000 Hz/s comparison velocity (8.8% ± 2.3%) than the 500 Hz/s comparison velocity (26.2% ± 4.5%, $p < 0.01$) or the 1500 Hz/s comparison velocity (16.6% ± 1.9%, $p < 0.01$). JNDs for the 500 Hz/s and 1500 Hz/s velocity conditions also differed significantly ($p < 0.05$). Neither the main effect of Task nor the Task x Comparison Velocity interaction reached significance [$F(1,38) = 1.19, \text{MSE} = 0.04, p = 0.28; F(2,76) = 0.22, \text{MSE} = 0.03, p = 0.80$, respectively].

Table 1. Points of Subject Equality (PSEs) for Experiment 1 (Standard Velocity = 1000 Hz/s), Experiment 2 (Standard Velocity = 500 Hz/s), and Experiment 3 (Standard Velocity = 1500 Hz/s).

<table>
<thead>
<tr>
<th>Standard Velocity</th>
<th>Comparison Velocity</th>
<th>500 Hz/s</th>
<th>1000 Hz/s</th>
<th>1500 Hz/s</th>
</tr>
</thead>
<tbody>
<tr>
<td>500 Hz/s</td>
<td>Time-change</td>
<td>5.7% (10.3%)</td>
<td>-7.2% (1.5%)</td>
<td>-67.9% (20.9%)</td>
</tr>
<tr>
<td></td>
<td>Pitch-change</td>
<td>-4.7% (10.5%)</td>
<td>2.6% (1.6%)</td>
<td>6.3% (21.5%)</td>
</tr>
<tr>
<td>1000 Hz/s</td>
<td>Time-change</td>
<td>33.4% (9.5%)</td>
<td>0.3% (1.4%)</td>
<td>-15.2% (19.5%)</td>
</tr>
<tr>
<td></td>
<td>Pitch-change</td>
<td>-45.5% (10.5%)</td>
<td>2.6% (1.6%)</td>
<td>6.3% (21.5%)</td>
</tr>
<tr>
<td>1500 Hz/s</td>
<td>Time-change</td>
<td>11.7% (11.2%)</td>
<td>5.5% (1.7%)</td>
<td>-4.7% (22.8%)</td>
</tr>
<tr>
<td></td>
<td>Pitch-change</td>
<td>-36.5% (11.2%)</td>
<td>-8.7% (1.7%)</td>
<td>4.2% (22.8%)</td>
</tr>
</tbody>
</table>

Table 2. Just Noticeable Difference thresholds (JNDs) for Experiment 1 (Standard Velocity = 1000 Hz/s), Experiment 2 (Standard Velocity = 500 Hz/s), and Experiment 3 (Standard Velocity = 1500 Hz/s).
Imputed velocity model

Finally, response proportions (‘longer’ and ‘more pitch change’) were fit with the imputed velocity model (Henry & McAuley, 2009; Henry, et al., 2009). Values of $w$ quantified the degree to which listeners relied on expectations for time or pitch change based on an assumption of constant velocity across the standard-comparison glide pair; smaller $w$ values correspond to larger degrees of perceptual distortion. For the time-change and pitch-change tasks, $w$ values were 0.74 (± 0.03) and 0.69 (± 0.04), respectively. These values did not differ significantly from each other, $t(38) = 0.97, p = 0.34$; that is, velocity contributed equally to perceived pitch and duration.

Summary

To summarize, the results of Experiment 1 were consistent with predictions derived from the auditory motion hypothesis. Comparison duration was overestimated for the relatively fast comparison velocity (1500 Hz/s) and underestimated for the relatively slow comparison velocity (500 Hz/s). Pitch change was underestimated for the relatively fast comparison velocity (1500 Hz/s) and overestimated for the relatively slow comparison velocity (500 Hz/s). No significant perceptual distortions for either task were apparent when the standard and comparison velocities were matched (1000 Hz/s). The imputed velocity model indicated a relatively large degree of perceptual distortion that did not differ between the time-change and pitch-change tasks.

One assumption made in interpreting the results of Experiment 1 was that listeners used the constant velocity of the standard glide to generate expectations about the duration or pitch change of the comparison glide, leading to perceptual distortions. However, it is also possible that the listener ignored the standard, and instead relied only on the relationship of the comparison velocities to each other, using an average comparison velocity as the constant-
velocity referent. In this case, varying the velocity of the standard glide should not affect perceived duration or pitch change. Experiments 2 and 3 directly tested this possibility. In Experiment 2, standard velocity was 500 Hz/s, and in Experiment 3, standard velocity was 1500 Hz/s. Thus, in Experiment 2, the standard and comparison glides matched for the slowest comparison velocity value (i.e., 500 Hz/s), and the remaining comparison velocity conditions (1000 Hz/s, 1500 Hz/s) were relatively fast. Based on the auditory motion hypothesis, it was expected that listeners would overestimate duration and underestimate pitch change for both the 1000 Hz/s and 1500 Hz/s comparison velocities, while duration and pitch change were expected to be judged relatively accurately for the 500 Hz/s comparison. In Experiment 3, the standard and comparison glides matched for the fastest comparison velocity value (i.e., 1500 Hz/s), and the remaining comparison velocities were relatively slow. Thus, based on the auditory motion hypothesis, listeners were expected to underestimate duration and overestimate pitch change for both the 500 Hz/s and 1000 Hz/s comparisons, while duration and pitch change were expected to be judged accurately for the 1500 Hz/s comparison. If on the other hand, if listeners did not make use of the standard, and instead generated expectations for comparison duration and pitch change based on an average of comparison velocities, the results of Experiments 2 and 3 would be expected to be identical to the results of Experiment 1.

Experiment 2

Method

Design

The design and task were the same as Experiment 1.

Participants

Thirty-seven individuals (n = 25, female) from a large Midwestern university participated
in exchange for course credit in an introductory psychology class. Listeners self reported normal hearing and had a range of formal musical training (0 – 12 years, $M = 3.0$, $SD = 3.3$). Listeners were randomly assigned to complete either the time-change task ($n = 19$) or the pitch-change task ($n = 18$).

*Stimuli and apparatus*

Stimuli were the same as in Experiment 1, with the following exceptions. Standard glides ascended at a velocity of 500 Hz/s. For the time-change task (Figure 7A), standard glides were 1000 ms in duration and comparison durations were centered on 1000 ms ($\pm 5\%$, $10\%$, $15\%$, $20\%$); standard pitch change was thus 500 Hz. For the pitch-change task (Figure 7B), standard pitch change was 1000 Hz and comparison frequency change was centered on 1000 Hz (1000 Hz $\pm 5\%$, $10\%$, $15\%$, $20\%$); standard duration was thus 2000 ms in duration.

Stimulus generation, stimulus presentation, and response collection were the same as in Experiment 1.

*Procedure*

The procedure was identical to Experiment 1.

*Results*

Figure 8 shows proportions of ‘longer’ responses (time-change task, Panel A) and ‘more pitch change’ responses (pitch-change task, Panel B) as a function of Comparison Level for each of the three comparison velocities (500 Hz/s, 1000 Hz/s, 1500 Hz/s). The current results are consistent with the idea that listeners made use of the standard velocity when generating expectations about the duration and pitch change of the comparison, supporting the auditory motion hypothesis. For the time-change task, perceived duration was relatively undistorted when comparison glide velocity was 500 Hz/s, that is, when the comparison velocity matched the
Figure 7. Stimulus diagram for Experiment 2. Standard glide velocity was 500 Hz/s. (A) For the time-change task, standard duration was 1000 ms and standard pitch change was 500 Hz. Comparison glides varied around 1000 ms (ΔT = ± 5%, 10%, 15%, or 20%), and listeners judged whether the comparison was ‘shorter’ or ‘longer’ than the standard. (B) For the pitch-change task, standard pitch change was 1000 Hz (ΔP = ± 5%, 10%, 15%, or 20%), and standard duration was 2000 ms. Listeners judged whether the comparison changed ‘more’ or ‘less’ in pitch than the standard.
Figure 8. Experiment 2. (A) Proportions of ‘longer’ responses and (B) proportions of ‘more pitch change’ responses as a function of Comparison Level for the time change and pitch-change tasks, respectively. When standard and comparison velocities were the same (500 Hz/s), listeners judged duration and pitch change relatively accurately. When the comparison velocity was relatively fast (1000 Hz/s, 1500 Hz/s), duration was overestimated and pitch change was underestimated.
standard velocity. Moreover, as predicted by the auditory motion hypothesis, comparison duration was overestimated when comparison glide velocity was relatively fast (1000 Hz/s, 1500 Hz/s). For the pitch-change task, perceived pitch change was relatively undistorted when the comparison glide and standard glide were presented at the same velocity (i.e., 500 Hz/s). Comparison pitch change was underestimated when comparison glide velocity was relatively fast (1000 Hz/s, 1500 Hz/s).

A 2 (Task) x 3 (Comparison Velocity) x 8 (Comparison Level) mixed-measures ANOVA on response proportions revealed a main effect of Task, $F(1,35) = 18.98$, $MSE = 0.15$, $p < 0.001$, $\eta^2_p = 0.35$, a main effect of Comparison Velocity, $F(2,70) = 5.55$, $MSE = 0.07$, $p < 0.01$, $\eta^2_p = 0.14$, a main effect of Comparison Level, $F(7,245) = 328.72$, $MSE = 0.04$, $p < 0.001$, $\eta^2_p = 0.90$, a Task x Comparison Velocity interaction, $F(2,70) = 99.10$, $MSE = 0.07$, $p < 0.001$, $\eta^2_p = 0.74$, a Task x Comparison Level interaction, $F(7,245) = 2.66$, $MSE = 0.04$, $p < 0.05$, $\eta^2_p = 0.07$, a Comparison Velocity x Comparison Level interaction, $F(14,490) = 6.85$, $MSE = 0.02$, $p < 0.001$, $\eta^2_p = 0.16$, and critically, a three-way Task x Comparison Velocity x Comparison Level interaction, $F(14,490) = 9.50$, $MSE = 0.02$, $p < 0.001$, $\eta^2_p = 0.21$. As in Experiment 1, the main effect of Comparison Level confirmed that proportions of ‘longer’ and ‘more pitch change’ responses increased with comparison duration and pitch change, respectively. The main effect of Task was due to overall more ‘longer’ responses ($M = 0.59 \pm 0.02$) than ‘more pitch change’ responses ($M = 0.48 \pm 0.02$), and the main effect of Comparison Velocity was due to decreasing proportions of ‘longer’ and ‘more pitch change’ responses with decreasing comparison velocity, $F_{lin} = 6.94$, $MSE = 0.11$, $p = 0.01$. This latter effect is difficult to interpret in light of the significant Task x Comparison Velocity interaction. Moreover, the two-way interactions must be interpreted in the context of the three-way interaction, which is consistent with the pattern of
results predicted by the auditory motion hypothesis (Figure 8). Specifically, comparison duration was overestimated and comparison pitch change was underestimated when comparison glide velocity was relatively fast (1000 Hz/s, 1500 Hz/s).

A 2 (Task) x 3 (Comparison Velocity) mixed-measures ANOVA on PSE failed to reveal any main effects or interactions (all $p$s > 0.11). However, inspection of Table 1 indicates that this was perhaps due to a large amount of variability and a notably small effect overall for the pitch-change task. Despite the lack of significance, the overall pattern of results was in line with predictions made by the auditory motion hypothesis; specifically, duration was overestimated and pitch change was underestimated for both relatively fast comparison velocities (1000 Hz/s, 1500 Hz/s). However, in contrast to the results of Experiment 1, PSEs for the 500 Hz/s (matched) comparison velocity were significantly different from 0 for both the time-change task ($PSE = 5.7\% \pm 1.1\%, t(18) = 5.00, p < 0.001$) and the pitch-change task ($PSE = -4.7\% \pm 1.0\%, t(17) = -4.70, p < 0.001$). Thus, small but non-zero perceptual distortions obtained when the standard and comparison velocities were equal.

A parallel analysis on JNDs revealed no main effects or interactions ($p$s > 0.20). Similar to the analysis on PSE, this is likely due partially to high variability and relatively small perceptual distortions for the pitch-change task (see Table 2). JNDs tended to be smallest for the 500 Hz/s comparison velocity; that is, judgments about duration and pitch change were most accurate when the standard and comparison glides were presented at the same velocity.

Imputed velocity model

Finally, proportions of ‘longer’ and ‘more pitch change’ responses were fit with the imputed velocity model to yield $w$ values that quantified the degree of perceptual distortion for time-change and pitch-change tasks; smaller values of $w$ correspond to larger perceptual
distortions. Values of $w$ for the time-change task and pitch-change tasks were 0.88 (± 0.06) and 0.89 (± 0.09), respectively. These values did not differ significantly from each other, $t(35) = -0.73, p = 0.47$; thus velocity contributed approximately equally to perceived pitch and duration. The velocity effects in the current experiment were relatively weak in comparison to the results of Experiment 1.

**Summary**

Experiment 2 addressed the possibility that the perceptual distortions observed in Experiment 1 did not reflect listeners’ generation of expectations about the comparison based on standard velocity, but instead indicated that listeners used an average of the comparison velocities to estimate comparison duration or pitch change. The results of Experiment 2 supported the auditory motion hypothesis, and the hypothesis that listeners made use of the constant-velocity referent provided by the standard in their judgments about the comparison. Specifically, comparison duration was overestimated and comparison pitch change was underestimated for the relatively fast comparison glide velocities (1000 Hz/s, 1500 Hz/s); however, the effects were overall weaker than in Experiment 1. Moreover, PSEs for both tasks were nonzero for the 1000 Hz/s comparison velocity (see Table 1), highlighting an important difference between the results of the current experiment and Experiment 1. That is, if listeners used as a velocity referent the average comparison velocity, PSEs for the 1000 Hz/s comparison velocity should have been close to 0, as in Experiment 1. However, non-zero PSEs for the 1000 Hz/s comparison indicate perceptual distortion, notably in the direction predicted by the auditory motion hypothesis under the assumption that listeners relied on the standard as the velocity referent.

Experiment 3 further tested the hypothesis that the distortions observed in Experiments 1
and 2 were based on use of the mean of comparison velocities rather than the standard velocity as the constant-velocity referent. In Experiment 3, the standard velocity was 1500 Hz/s. Thus for both relatively slow comparison velocities (500 Hz/s, 1000 Hz/s), listeners were expected to underestimate comparison duration and overestimate comparison pitch change. As before, listeners were expected to most accurately judge comparison duration and pitch change when the comparison velocity matched the standard (i.e., 1500 Hz/s).

Experiment 3

Method

Design

The design and task were the same as Experiments 1 and 2.

Participants

Thirty-two individuals (n = 23, female) from a large Midwestern university participated in exchange for course credit in an introductory psychology class. Listeners self reported normal hearing and had a range of formal musical training (0 – 15 years, M = 4.5, SD = 3.7). Listeners were randomly assigned to complete either the time-change task (n = 16) or the pitch-change task (n = 16).

Stimuli and apparatus

Stimuli were the same as in the previous experiments, with the following exceptions. Standard glides ascended at a velocity of 1500 Hz/s. For the time-change task (Figure 9A), standard glides were 1000 ms in duration and comparison duration was centered on 1000 ms (± 5%, 10%, 15%, 20%); standard pitch change was thus 1500 Hz. For the pitch-change task (Figure 9B), standard pitch change was 1000 Hz, and comparison pitch change was varied around 1000 Hz (± 5%, 10%, 15%, 20%); comparison duration was thus 667 ms. Stimulus
(A) Time-change task

Standard Glide: 1500 Hz/s

Comparison Glide: 500 Hz/s, 1000 Hz/s, 1500 Hz/s

(B) Pitch-change task

Standard Glide: 1500 Hz/s

Comparison Glide: 500 Hz/s, 1000 Hz/s, 1500 Hz/s

Figure 9. Stimulus diagram for Experiment 3. Standard glide velocity was 1500 Hz/s. (A) For the time-change task, standard duration was 1000 ms and standard pitch change was 1500 Hz. Comparison glides varied around 1000 ms (ΔT = ± 5%, 10%, 15%, or 20%), and listeners judged whether the comparison was ‘shorter’ or ‘longer’ than the standard. (B) For the pitch-change task, standard pitch change was 1000 Hz, and standard duration was 667 ms. Comparison glides varied around 1000 Hz (ΔP = ± 5%, 10%, 15%, or 20%), and listeners judged whether the comparison changed ‘more’ or ‘less’ in pitch than the standard.
generation, stimulus presentation, and response collection were the same as in Experiments 1 and 2.

Procedure

The procedure was identical to Experiments 1 and 2.

Results

Figure 10 shows proportions of ‘longer’ responses (time-change task, Panel A) and ‘more pitch change’ responses (pitch-change task, Panel B) as a function of Comparison Level for each of the three comparison velocities (500 Hz/s, 1000 Hz/s, 1500 Hz/s). Results were overall consistent with the auditory motion hypothesis. For the time-change task, perceived duration was relatively undistorted when comparison glide velocity was 1500 Hz/s, that is, when the comparison velocity matched the standard velocity. Moreover, comparison duration was underestimated for both relatively slow comparison velocities (500 Hz/s, 1000 Hz/s). For the pitch-change task, perceived pitch change was relatively undistorted when the comparison glide and standard glide were presented at the same velocity, and pitch change was overestimated for both relatively slow comparison velocities (500 Hz/s, 1000 Hz/s).

Supporting this interpretation, a 2 (Task) x 3 (Comparison Velocity) x 8 (Comparison Level) mixed-measures ANOVA on response proportions revealed a main effect of Task, $F(1,30) = 33.52$, $MSE = 0.16$, $p < 0.001$, $\eta^2_p = 0.53$, a main effect of Comparison Level, $F(7,210) = 259.62$, $MSE = 0.04$, $p < 0.001$, $\eta^2_p = 0.90$, a Task x Comparison Velocity interaction, $F(2,60) = 51.09$, $MSE = 0.13$, $p < 0.001$, $\eta^2_p = 0.63$, a Comparison Velocity x Comparison Level interaction, $F(14,420) = 9.41$, $MSE = 0.02$, $p < 0.001$, $\eta^2_p = 0.24$, and critically, a three-way Task x Comparison Velocity x Comparison Level interaction, $F(14,420) = 9.50$, $MSE = 0.02$, $p < 0.001$, $\eta^2_p = 0.24$. The main effect of Comparison Level confirms that proportions of ‘longer’ and
Figure 10. Experiment 3. (A) Proportions of ‘longer’ responses and (B) proportions of ‘more pitch change’ responses as a function of Comparison Level for the time change and pitch-change tasks, respectively. When the standard and comparison velocities were the same (1500 Hz/s), listeners judged duration and pitch change relatively accurately. When comparison velocity was relatively slow (500 Hz/s, 1000 Hz/s), duration was underestimated and pitch change was overestimated.
‘more pitch change’ responses increased with comparison duration and pitch change, respectively. The main effect of Task was driven by a larger proportion of ‘more pitch change’ responses ($M = 0.62 \pm 0.02$) than ‘longer’ responses ($M = 0.45 \pm 0.02$); however, the Task effect was qualified by the Task x Comparison Velocity interaction. Moreover, the two-way interactions must be interpreted in the context of the three-way interaction, which is consistent with the pattern of results predicted by the auditory motion hypothesis; specifically, comparison duration was underestimated and comparison pitch change was overestimated when comparison glide velocity was relatively slow (500 Hz/s, 1000 Hz/s).

A 2 (Task) x 3 (Comparison Velocity) mixed-measures ANOVA on PSE supported the analysis of response proportions. Specifically, the ANOVA revealed a main effect of Task, $F(1,30) = 8.30, MSE = 0.09, p < 0.01, \eta^2_p = 0.22$, and a Task x Comparison Velocity interaction, $F(2,60) = 6.91, MSE = 0.10, p < 0.01, \eta^2_p = 0.19$. The main effect of Comparison Velocity was not significant ($p = 0.24$). The main effect of Task indicated that, overall, PSEs tended to be somewhat larger for the pitch-change task ($PSE = -13.7\% \pm 4.4\%$) than for the time-change task ($PSE = 4.1\% \pm 4.4\%$). However, the main effect of Task was qualified by the Task x Comparison Velocity interaction (Table 1); as predicted by the auditory motion hypothesis, duration was underestimated and pitch change was overestimated when comparison glide velocity was relatively slow (500 Hz/s, 1000 Hz/s). Single-sample $t$-tests indicated that mean PSE for the 1500 Hz/s comparison velocity was slightly different from 0 in the time-change task, $t(15) = -3.34, p < 0.01$, but not significantly different from 0 in the pitch-change task, $t(15) = 1.59, p = 0.13$.

A parallel analysis on JND revealed only a main effect of Comparison Velocity, $F(2,60) = 3.66, MSE = 0.03, p < 0.05, \eta^2_p = 0.11$ (see Table 2). Discrimination thresholds were higher for
the 500 Hz/s comparison glide velocity ($JND = 20.0\% \pm 4.8\%$) than for the 1500 Hz/s velocity ($JND = 10.0\% \pm 0.6\%$), $p = 0.05$. The difference between JNDs for the 500 Hz/s and 1000 Hz/s ($JND = 11.1\% \pm 0.8\%$) glide velocities also trended towards significance, $p = 0.08$. JNDs did not differ significantly for the 1000 Hz/s and 1500 Hz/s comparison velocity conditions, $p = 0.13$.

**Imputed velocity model**

Next, as in the previous experiments, proportions of ‘longer’ and ‘more pitch change’ responses were fit with the imputed velocity to yield values of $w$ that quantify the degree to which perceived duration and pitch change were distorted based on an assumption of constant velocity across the standard-comparison pair; smaller values of $w$ correspond to larger perceptual distortions. Values of $w$ for the time-change and pitch-change tasks were 0.79 ($\pm 0.03$) and 0.77 ($\pm 0.07$), respectively. These values were not significantly different from each other, $t(30) = 0.28$, $p = 0.78$; thus velocity contributed equally to perceived pitch and duration. Values of $w$ obtained in the current experiment were similar to Experiment 1.

**Comparison of results across standard velocities**

In order to compare results across the three experiments, proportions of ‘longer’ (time-change task) and ‘more pitch change’ (pitch-change task) responses were entered into a 2 (Task) x 3 (Standard Velocity) x 3 (Comparison Velocity) x 8 (Comparison Level) mixed-measures ANOVA to test the specific prediction that the four-way interaction would reach significance. A significant four-way interaction, $F(28,1442) = 3.29$, $MSE = 0.02$, $p < 0.001$, $\eta^2_p = 0.06$, confirmed three different patterns of results across experiments, and thus supported the predictions made by the auditory motion hypothesis. Specifically, in Experiment 1, when standard glide velocity was 1000 Hz/s, comparison duration was overestimated and comparison pitch change was underestimated when comparison glide velocity was relatively fast (1500 Hz/s).
Hz/s). Moreover, comparison duration was underestimated and comparison pitch change was
overestimated when comparison glide velocity was relatively slow (500 Hz/s). In Experiment 2,
when standard glide velocity was 500 Hz/s, comparison duration was overestimated and pitch
change was underestimated for both relatively fast comparison velocities (1000 Hz/s, 1500 Hz/s).
Finally, in Experiment 3, when the standard velocity was equal to 1500 Hz/s, duration was
underestimated and pitch change was overestimated for both relatively slow comparison
velocities (1000 Hz/s, 1500 Hz/s).

Values of $w$ derived from the imputed velocity model were also examined in a 2 (Task) x
3 (Standard Velocity) between-subjects ANOVA to assess effects of Standard Velocity on the
magnitude of perceptual distortion. The main effect of Standard Velocity reached significance,
$F(2,108) = 13.01, MSE = 0.03, p < 0.001, \eta^2_p = 0.19$, indicating higher $w$ values for the 500 Hz/s
standard velocity ($w = 0.88 \pm 0.03$) than for either the 1000 Hz/s ($w = 0.70 \pm 0.04$) or 1500 Hz/s
standard velocities ($w = 0.77 \pm 0.03$), $ps < 0.05$, which were not different from each other, $p =
0.13$, according to Tukey’s HSD post-hoc test. In line with previous work on discrete sequences
by Henry and colleagues (Henry & McAuley, 2009; Henry, et al., 2009), values of $w$ were
largest for the slowest Standard Velocity.

Finally, an exploratory analysis correlated $w$ values with years of formal musical training
to assess the potential for influences of individual differences on pitch-time interdependence. In
this regard, it has been previously suggested that formal musical training increases listeners’
ability to selectively attend to pitch or time (Monahan, Kendall, & Carterette, 1987; Monahan &
Carterette, 1985; Neuhaus & Knösche, 2008; Pitt & Monahan, 1987). For the time-change task, a
significant negative correlation obtained between years of musical training and $w, r(110) = -0.31,
p < 0.05$, whereas a strong positive correlation was observed for the pitch-change task, $r(100) =$
0.45, \( p < 0.01 \). These correlations can be interpreted to mean that more years of musical training was associated with better ability to ignore time while attending to pitch, but less success ignoring pitch while attending to time.

**Summary**

To summarize the results of Experiment 3, in line with the auditory motion hypothesis, comparison duration was underestimated and comparison pitch change was overestimated for both relatively slow comparison velocity conditions (500 Hz/s, 1000 Hz/s). A non-zero mean PSE for the 1500 Hz/s comparison velocity in the time-change task evidenced a small distortion in perceived duration when the standard and comparison velocities matched; however, mean PSE for the 1500 Hz/s velocity did not differ from 0 for the pitch-change task. A comparison across experiments confirmed different patterns of response proportions for each Standard Velocity, and thus supported the auditory motion hypothesis. Moreover, a comparison of \( w \) values across experiments indicated slightly higher values of \( w \), and thus smaller perceptual distortions, for the slowest Standard Velocity condition. Finally, a correlational analysis revealed that musical training was associated with ability to ignore time while judging pitch, but less success ignoring pitch when judging time.

**General Discussion**

Three experiments provided an extension of the auditory motion hypothesis to sounds moving continuously in pitch space. Listeners judged the duration (i.e., time change) or pitch change of a comparison tone glide with variable velocity (500 Hz/s, 1000 Hz/s, 1500 Hz/s) relative to the fixed duration or pitch change of a standard tone glide with constant velocity. It was expected that listeners would assume constant velocity across the standard-comparison pair, and generate expectations about comparison duration or pitch change based on assumed constant
Auditory Motion for Continuous and Discrete Sounds

velocity.

In Experiment 1, standard glide velocity was 1000 Hz/s. When comparison glide velocity was faster than standard velocity (i.e., 1500 Hz/s), comparison duration was overestimated and comparison pitch change was underestimated. When comparison glide velocity was slower than standard velocity (i.e., 500 Hz/s), comparison duration was underestimated and comparison pitch change was overestimated. In Experiment 2, standard glide velocity was 500 Hz/s. When comparison glide velocity was faster than the standard (1000 Hz/s, 1500 Hz/s), comparison duration was overestimated and comparison pitch change was underestimated. Finally, in Experiment 3, standard glide velocity was 1500 Hz/s. When comparison glide velocity was slower than standard velocity (500 Hz/s, 1000 Hz/s), comparison duration was underestimated and comparison pitch change was overestimated. In all three experiments, when standard velocity and comparison velocity were the same, perceived time and pitch change were relatively undistorted. Analyses on PSEs and JNDs supported these results, indicating that listeners most accurately judged duration and pitch change when standard and comparison velocities matched, but perception was distorted when standard and comparison velocities were different.

Overall, consistent with the auditory motion hypothesis, these results support the hypothesis that listeners made use of the standard velocity to generate expectations about the to-be-judged duration and pitch change of the comparison. An important possibility was ruled out by taking the results of the three experiments together. That is, it could be proposed that the effect observed in Experiment 1 was not at all related to the velocity of the standard, but was instead due to convergence of a remembered standard to the mean of the comparison velocities. In reminder tasks that involve the presentation of a fixed standard on each trial followed by a variable comparison, it has been shown that listeners sometimes ignore the standard altogether,
and instead compare the current value of the comparison to a remembered standard (e.g., Miller & McAuley, 2005). Additionally, when comparison values vary along a continuum, a convergence towards the mean value is often observed such that larger magnitude stimuli tend to be underestimated while smaller magnitude stimuli tend to be overestimated (Vierordt, 1868). Critically, if listeners ignored the standard and used as a velocity referent the mean comparison velocity value, the pattern of results across all three experiments should have been identical. Instead, when response proportions for all three experiments were compared, a four-way interaction was observed, the form of which was predicted by the auditory motion hypothesis.

In the current experiment, ‘shorter’ and ‘more pitch change’ responses were always associated with a left-side button press, while ‘longer’ and ‘less pitch change’ responses were always associated with a right-side button press. Thus, an alternative explanation for the current results comes from research on stimulus-response compatibility (SRC). Previously, SRC effects have been shown for both time (Ishihara, Keller, Rossetti, & Prinz, 2008) and pitch (Keller & Koch, 2006; Lidji, et al., 2007; Rusconi, et al., 2006) judgments. With respect to time, listeners tend to associate ‘early’ events with left-side responses and ‘late’ events with right-side responses. With respect to pitch, the preferred mapping is vertical, with high pitches associated with high buttons and low pitches with low buttons, although musicians tend to show a horizontal mapping of low pitches to left-side responses and high pitches to right-side responses. It is not unreasonable to propose that such mappings might exist for velocity and could potentially explain the current results. This explanation depends on mappings between standard-comparison velocity relationships and response buttons; specifically, slow comparison velocities (relative to the standard) would have to map to a left-side response, and fast comparison velocities (relative to the standard) would have to map to a right-side response. For the time-
change task then, relatively slow comparisons would map onto a ‘shorter’ response (leading to an underestimation of duration), and relatively fast comparisons would map onto a ‘longer’ response (leading to an overestimation of duration). For the pitch-change task, relatively slow comparisons would map to a ‘more pitch change’ response (leading to an overestimation of pitch change), and relatively fast comparisons would map to a ‘less pitch change’ response (leading to an underestimation of pitch change).

In order to rule out this possibility, Experiment 1 was replicated with one exception. The spatial positions of response buttons were reversed for both tasks, so that for the time-change task ‘shorter’ and ‘longer’ responses were made by pushing the right-side and left-side buttons, respectively, and for the pitch-change task, ‘less pitch change’ and ‘more pitch change’ responses were made by pushing the left-side and right-side buttons, respectively. Figure 11 shows results for this experiment, which involved thirty undergraduates from a large Midwestern university (n = 12, time-change task; n = 13, pitch-change task). If SRC effects explain the results of Experiments 1–3, then the current results (Figure 11) should be opposite those obtained in Experiment 1; that is, duration should have been underestimated and pitch-change overestimated for the relatively fast comparison velocity (1500 Hz/s), while duration should have been overestimated and pitch-change underestimated for the relatively slow comparison velocity (500 Hz/s). However, it is clear from inspection of the figure that the pattern is identical to that obtained in Experiment 1. Thus it is considered highly unlikely that SRC effects account for the result of Experiments 1–3.

In its current form, the auditory motion hypothesis does not distinguish between continuous and discrete auditory stimuli. In this regard, it is notable that the current research constitutes the first application of the auditory motion hypothesis to stimuli moving continuously
Figure 11. Results from a replication of Experiment 1 with the exception that the spatial location of response buttons was reversed. (A) Proportions of ‘longer’ responses and (B) proportions of ‘more pitch change’ responses as a function of Comparison Level for the time change and pitch-change tasks, respectively. As in Experiment 1, duration and pitch change were most accurately judged when standard and comparison velocities matched (1000 Hz/s). When comparison velocity was relatively fast (1500 Hz/s), duration was overestimated and pitch change was underestimated. When comparison velocity was relatively slow (500 Hz/s), duration was underestimated and pitch change was overestimated.
in pitch space. In Chapter 1, a study conducted by Henry and McAuley (in press) was reviewed that showed that listeners seemed to rely on different information when estimating velocity of continuous versus discrete auditory stimuli. For continuous tone glides, pitch-change and time-change information contributed approximately equally to estimates of perceived velocity. However, for discrete tone sequences, only time-change information contributed to velocity estimates, with little to no contribution of pitch-change information. These results highlight potentially important differences in velocity perception for continuous versus discrete auditory stimuli moving in pitch space. Another relevant finding from this study concerns estimates of total duration and total pitch change for isochronous sequences. When listeners separately estimated the total pitch change and time change of isochronous sequences, their responses were uninfluenced by variation along the to-be-ignored dimension. This latter result raises the possibility that listeners may be better able to selectively attend to the to-be-judged dimension in discrete auditory stimuli that vary in pitch and time.

Thus the next set of experiments considers the possible influence of assumed constant velocity on perceived duration and pitch change of discrete tone sequences in a paradigm matched to the experiments reported in this chapter. If continuous and discrete auditory stimuli are treated similarly by the listener in terms of pitch-time interdependence, then the results for discrete tone sequences should be the same as those for continuous tone glides. However, if listeners are able to selectively attend to time and pitch in discrete sequences, as is implied by the results of Henry and McAuley (in press, Experiment 3), then no effects of velocity should be observed. One final possibility, also suggested by the results of Henry and McAuley (in press, Experiment 1) is that listeners obligatorily attend to time in discrete stimuli. In this case, responses should depend on temporal characteristics of the standard and comparison sequences.
rather than on velocity. Specifically, listeners should perform the time-change task accurately, regardless of the velocities of the standard and comparison stimuli. However, for the pitch-change task, velocity should exert a strong influence on pitch-change judgments, in that total sequence duration is confounded with velocity.
CHAPTER 3: EXTENSION OF THE AUDITORY MOTION HYPOTHESIS TO SOUNDS
IMPLYING MOTION IN FREQUENCY SPACE

In the experiments reported in this chapter, listeners judged either the total duration (time-change task) or total pitch change (pitch-change task) of a variable-velocity comparison tone sequence relative to a constant-velocity standard tone sequence. For the time-change task, the standard duration was fixed at 750 ms from the onset of the first tone to the onset of the final tone, and the duration of the comparison varied around 750 ms (± 5%, 10%, 15%, 20%). For the pitch-change task, the total pitch change of the standard was fixed at 1000 Hz, and the total pitch change of the comparison varied around 1000 Hz (± 5%, 10% 15%, 20%). Across all experiments, the comparison velocity was variable and took on values of 667 Hz/s, 1333 Hz/s, and 2000 Hz/s. The standard velocity was varied across experiments. In Experiment 4, standard sequence velocity was 1333 Hz/s, in Experiment 5, standard sequence velocity was 667 Hz/s, and in Experiment 6, the standard sequence velocity was 2000 Hz/s.

There were several possible outcomes of Experiments 4 – 6. First, if motion in pitch space implied by discrete tone sequences was treated equivalently to continuous motion in pitch space, then the results of Experiments 4, 5, and 6 would be expected to replicate the results of Experiments 1, 2, and 3, respectively. However, there are reasons to expect that implied and actual motion in pitch space are not perfectly analogous. In particular, it has been suggested that velocity is perceived differently in continuous and discrete auditory stimuli, with perceived velocity for continuous stimuli predictable from joint pitch and time information, and perceived velocity for discrete stimuli predictable on the basis of temporal properties (Henry & McAuley, in press). Thus, a second possibility was that listeners would be unable to attend to pitch information in discrete tone sequences, and so both duration and pitch-change judgments would
reflect the timing characteristics of the sequences. Finally, a third possibility also suggested by Henry and McAuley (in press, Experiment 3) was that the discrete nature of the stimuli in the current experiments would allow listeners to better selectively attend to the time and pitch dimensions. In their Experiment 3, listeners were asked to make independent estimates of the total duration and total pitch change of discrete tone sequences; results indicated that listeners did so without being influenced by variation along the to-be-ignored dimension. On this view, listeners would not use assumed constant velocity to generate expectations about the duration or pitch change of the comparison relative to the standard. The result would be undistorted perception of time and pitch, regardless of the relationship of the standard and comparison velocities.

Experiment 4

Method

Design

The design was a 2 (Task: time-change, pitch-change) x 3 (Comparison Velocity: 667 Hz/s, 1333 Hz/s, 2000 Hz/s) x 8 (Comparison Level: -20%, -15%, -10%, -5%, +5%, +10%, +15%, +20%) mixed-factorial. Task was a between-subjects factor; Comparison Velocity and Comparison Level were within-subjects factors. Participants judged either the total duration (time-change task; Figure 12A) or the total pitch change (pitch-change task; Figure 12B) of a comparison tone sequence relative to a standard.

Participants

Thirty-four individuals (n = 32, female) from a large Midwestern university participated in exchange for course credit in an introductory psychology class. Listeners self reported normal hearing and had a range of formal musical training (0 – 11 years, $M = 3.4$ years, $SD = 3.6$ years).
(A) Time-change task

Standard Sequence: 1333 Hz/s

Comparison Sequence:
667 Hz/s, 1333 Hz/s, 2000 Hz/s

(B) Pitch-change task

Standard Glide: 1000 Hz/s

Comparison Glide:
500 Hz/s, 1000 Hz/s, 1500 Hz/s

Figure 12. Stimulus diagram for Experiment 4. Standard sequence velocity was 1333 Hz/s. For both the time change and pitch-change tasks, standard glide duration was 750 ms and standard glide pitch change was 1000 Hz. (A) For the time-change task, $\Delta T = \pm 5\%, 10\%, 15\%, \text{ or } 20\%$, and listeners judged whether the comparison was ‘shorter’ or ‘longer’ than the standard. (B) For the pitch-change task, $\Delta P = \pm 5\%, 10\%, 15\%, \text{ or } 20\%$, and listeners judged whether the comparison changed ‘more’ or ‘less’ in pitch than the standard.
Listeners were randomly assigned to complete either the time-change task (n = 19) or the pitch-change task (n = 15).

Stimuli and apparatus

Stimuli were isochronous four-tone sequences comprised of 100-ms pure tones; each tone was ramped over the first and last 5 ms to eliminate acoustic artifacts. Standard sequences ascended at a velocity of 1333 Hz/s, and comparison sequences ascended at one of three velocities (667 Hz/s, 1333 Hz/s, 2000 Hz/s). As in Experiments 1–3, velocity was specified in terms of linear frequency space, i.e., Hz/s. Standard sequence duration (defined from the onset of the first tone to the onset of the last tone) was 750 ms; this means that stimulus onset asynchrony (SOA) between individual sequences tones was 250 ms. Comparison duration was varied around 750 ms (± 5%, 10%, 15%, 20%). Importantly, varying the total duration of the comparison tone sequences resulted in changes to the SOA such that (when the standard and comparison velocity matched), the rate of comparison sequences was predictive of total duration. The total pitch change of the standard sequence was 1000 Hz. Comparison pitch change was varied around 1000 Hz (± 5%, 10%, 15%, 20%). The starting pitch of the standard and comparison sequences was randomized from trial to trial, taking on one of three values (476 Hz, 600 Hz, 756 Hz);

Stimuli were generated using MATLAB software (The Mathworks, Inc.). Stimulus generation and response collection were controlled using E-Prime 2.0.8.73 software (Psychology Software Tools, Inc.) running on Dell Optiplex computers. Auditory stimuli were presented at a comfortable listening level (~70 dB) over Sennheiser HD 280 Pro headphones (Old Lyme, CT). Responses were made using a serial response box. ‘Shorter’ and ‘more pitch change’ responses were made by pushing a left-side button, and ‘longer’ and ‘less pitch change’ responses were
made by pushing a right-side button.

Procedure

Participants heard recorded instructions, then completed a short 18-trial training block with corrective feedback. During training, listeners heard only comparison levels of ± 20%, completing one trial per Comparison Velocity at each starting pitch value. Listeners then completed two experimental blocks with no feedback. On each trial, participants heard a standard-comparison tone sequence pair, and judged whether the comparison was ‘shorter’ or ‘longer’ than the standard (time-change task) or whether the comparison changed ‘more’ or ‘less’ in pitch than the standard (pitch-change task). The standard sequence ascended at a velocity of 1333 Hz /s from one of three starting pitches (476 Hz, 600 Hz, 756 Hz). The comparison sequence ascended at one of three velocities (667 Hz /s, 1333 Hz /s, 2000 Hz/s) from one of the same three starting pitches. For the time-change task, the comparison duration was varied around 750 ms (± 5%, 10%, 15%, 20%), and for the pitch-change task, the comparison pitch change was varied around 1000 Hz (± 5%, 10%, 15%, 20%).

In each experimental block, listeners provided three responses to each combination of Comparison Velocity and Comparison Level, for a total of 72 trials. Overall, listeners completed a total of 144 trials in two experimental blocks; thus 6 observations were obtained per Comparison Velocity x Comparison Level combination. Following completion of the experiment, all listeners completed several surveys assessing aspects of musical and linguistic background, effort, understanding, attention to the task, and perceived task difficulty. The experiment lasted approximately 30 minutes.

Results

Figure 13 shows proportions of ‘longer’ (time-change task, Panel A) and ‘more pitch
Figure 13. Experiment 4. (A) Proportions of ‘longer’ responses and (B) ‘more pitch change’ responses as a function of Comparison Level for the time-change and pitch-change tasks, respectively. There were no systematic effects of Comparison Velocity.
change’ responses (pitch-change task, Panel B) as a function of Comparison Level for each of the three comparison velocities (667 Hz/s, 1333 Hz/s, 2000 Hz/s). For the both the time-change and pitch-change tasks, proportions of ‘longer’ and ‘more pitch change’ responses increased as a function of Comparison Level. However, in contrast to the results with continuous glide tones (Experiments 1 – 3), there were no systematic effects of Comparison Velocity.

A 2 (Task) x 3 (Comparison Velocity) x 8 (Comparison Level) mixed-measures ANOVA on response proportions revealed a main effect of Comparison Level, $F(7,224) = 315.59, MSE = 0.05, p < 0.001, \eta^2_p = 0.91$, confirming that proportions of ‘longer’ and ‘more pitch change’ responses increased with comparison duration and pitch change, respectively. Moreover, the main effect of Comparison Velocity, $F(2,64) = 6.60, MSE = 0.06, p < 0.01, \eta^2_p = 0.17$, reached significance. Overall, there were more ‘longer’ and ‘more pitch change’ responses to the 500 Hz/s comparison velocity ($M = 0.56 \pm 0.02$), than either the 1000 Hz/s ($M = 0.50 \pm 0.01$), $p < 0.01$, or the 1500 Hz/s comparison velocity ($M = 0.50 \pm 0.02$), $p < 0.01$. The Task x Comparison Velocity interaction was significant, $F(2,64) = 3.53, MSE = 0.06, p < 0.05, \eta^2_p = 0.10$, and the Comparison Velocity x Comparison Level interaction was marginally significant, $F(14,448) = 1.57, MSE = 0.02, p = 0.08, \eta^2_p = 0.03$. The Task x Comparison Velocity interaction was driven by a slight increase in proportions of ‘more pitch change’ relative to ‘longer’ responses for negative comparison levels. The Comparison Velocity x Comparison Level interaction was due to a leftward shift of the psychometric function for the 667 Hz/s comparison velocity relative to the 1333 Hz/s and 2000 Hz/s comparison velocities that was only apparent in the middle of the Comparison Level range; that is, response proportions converged at the extreme Comparison Levels.

A 2 (Task) x 3 (Comparison Velocity) mixed-measures ANOVA on PSE supported the
analysis of response proportions (see Table 3). A main effect of Comparison Velocity, $F(2.64) = 5.90$, $MSE = 23.74$, $p < 0.01$, $\eta^2_p = 0.16$, indicated that duration and pitch change were overestimated somewhat for the 500 Hz/s comparison velocity condition ($M = -3.41 \pm 1.09$) relative to the 1000 Hz/s ($M = 0.37 \pm 0.59$), $p < 0.01$, and 1500 Hz/s ($M = -0.16 \pm 0.96$), $p < 0.05$, comparison velocities, which did not differ from each other, $p = 0.84$. The Task x Comparison Velocity interaction was also marginally significant, $F(2,64) = 2.23$, $MSE = 23.74$, $p = 0.10$, $\eta^2_p = 0.07$; however, PSEs were not different for the time-change versus the pitch-change task for any value of Comparison Velocity ($ps > 0.10$).

<table>
<thead>
<tr>
<th>Standard Velocity</th>
<th>500 Hz/s</th>
<th>1000 Hz/s</th>
<th>1500 Hz/s</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time-change</td>
<td>-0.7% (1.1%)</td>
<td>0.3% (0.7%)</td>
<td>1.1% (1.4%)</td>
</tr>
<tr>
<td>Pitch-change</td>
<td>-0.7% (1.1%)</td>
<td>1.0% (0.7%)</td>
<td>1.0% (1.4%)</td>
</tr>
<tr>
<td>Time-change</td>
<td>-2.7% (1.0%)</td>
<td>-0.6% (0.7%)</td>
<td>1.4% (1.3%)</td>
</tr>
<tr>
<td>Pitch-change</td>
<td>-4.1% (1.2%)</td>
<td>1.4% (0.8%)</td>
<td>-1.7% (1.5%)</td>
</tr>
<tr>
<td>Time-change</td>
<td>-2.4% (1.1%)</td>
<td>-1.4% (0.7%)</td>
<td>1.0% (1.4%)</td>
</tr>
<tr>
<td>Pitch-change</td>
<td>-1.9% (1.1%)</td>
<td>0.5% (0.8%)</td>
<td>-1.1% (1.5%)</td>
</tr>
</tbody>
</table>

Table 3. Points of Subject Equality (PSEs) for Experiment 4 (Standard Velocity = 1333 Hz/s), Experiment 5 (Standard Velocity = 667 Hz/s), and Experiment 6 (Standard Velocity = 2000 Hz/s).

Finally, a 2 (Task) x 3 (Comparison Velocity) mixed-measures ANOVA for JNDS revealed no main effects or interactions ($ps \geq 0.52$). Discrimination thresholds for time judgments and pitch-change judgments were similar for all three Comparison Velocity conditions (see Table 4).
Comparison Velocity

<table>
<thead>
<tr>
<th>Task</th>
<th>500 Hz/s</th>
<th>1000 Hz/s</th>
<th>1500 Hz/s</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time-change</td>
<td>9.3% (0.9%)</td>
<td>11.7% (1.4%)</td>
<td>9.5% (1.0%)</td>
</tr>
<tr>
<td>Pitch-change</td>
<td>10.0% (0.9%)</td>
<td>10.4% (1.4%)</td>
<td>12.0% (1.0%)</td>
</tr>
<tr>
<td>Time-change</td>
<td>10.0% (0.9%)</td>
<td>9.5% (1.4%)</td>
<td>9.0% (1.0%)</td>
</tr>
<tr>
<td>Pitch-change</td>
<td>10.4% (1.0%)</td>
<td>9.5% (1.5%)</td>
<td>10.6% (1.1%)</td>
</tr>
<tr>
<td>Time-change</td>
<td>8.9% (0.9%)</td>
<td>9.0% (1.4%)</td>
<td>8.7% (1.0%)</td>
</tr>
<tr>
<td>Pitch-change</td>
<td>9.2% (1.0%)</td>
<td>9.4% (1.5%)</td>
<td>9.8% (1.1%)</td>
</tr>
</tbody>
</table>

Table 4. Just Noticeable Difference thresholds (JNDs) for Experiment 4 (Standard Velocity = 1333 Hz/s), Experiment 5 (Standard Velocity = 667 Hz/s), and Experiment 6 (Standard Velocity = 2000 Hz/s).

**Imputed velocity model**

As in Experiments 1 – 3, response proportions (‘longer’ and ‘more pitch change’) were fit with the imputed velocity model (Henry & McAuley, 2009; Henry, et al., 2009). Values of $w$ close to 1 indicate little contribution of velocity information, while decreasing $w$ values indicate increasing degrees of perceptual interdependence. Values of $w$ for the time-change tasks and pitch-change task were $w = 0.99$ (± 0.03) and $w = 0.97$ (± 0.04), respectively. The magnitude of perceptual distortions was overall very small; the difference between $w$ values for the time judgment and pitch-change tasks was marginally significant, $t(32) = 1.98$, $p = 0.057$. The slightly smaller $w$ value for the pitch-change task was due to a subtle leftward shift of the psychometric curve for the 677 Hz/s comparison velocity condition.

**Summary**

To summarize the results of Experiment 4, listeners overall performed the task as expected, responding ‘longer’ when the comparison sequence was longer in duration than the standard, and responding ‘more pitch change’ when the comparison changed more in pitch than the standard. However, differences dependent on Task and Comparison Velocity were subtle and
not systematic. Specifically, in contrast to the results of Experiment 1, there was no systematic
over- or underestimation of comparison duration or pitch change based on the relationship of the
Comparison Velocity to the Standard Velocity. Fitting the imputed velocity hypothesis to the
data supported this conclusion; \( w \) values for both the time judgment and pitch-change tasks were
close to 1.

The results of the current study are in contrast to the results obtained in Experiment 1,
and indicate that listeners were capable of attending selectively to time and pitch in discrete
sequences that implied motion in pitch space; that is, there was no evidence that listeners made
use of velocity information to generate expectations about comparison duration and pitch
change. This is consistent with the results reported by Henry and McAuley (in press) showing
that, despite the apparent inability of listeners to ignore timing information in discrete sequences
when estimating velocity, independent judgments of total duration and total pitch change were
quite accurate and unaffected by variation on the to-be-ignored dimension. Thus, it seems that
discrete sequences afford the listener the ability to attend selectively to the dimensions of time
and pitch. This hypothesis was tested further in Experiments 5 and 6, where the standard velocity
was varied and took on values of 667 Hz/s and 2000 Hz/s, respectively. Based on the results of
Experiment 4, listeners were expected to accurately judge total duration and pitch change of
comparison tone sequences relative to the standard, and no systematic velocity effects were
expected.

Experiment 5

Method

Design

The design and task were identical to Experiment 4.
Participants

Thirty-six individuals (n = 27, female) from a large Midwestern university participated in exchange for course credit in an introductory psychology class. Listeners self reported normal hearing and had a range of formal musical training (0 – 14 years, $M = 4.2$ years, $SD = 4.4$ years). Listeners were randomly assigned to complete either the time-change task (n = 18) or the pitch-change task (n = 18).

Stimuli and apparatus

Stimuli were the same as in Experiment 4, with the following exceptions. Standard sequences ascended at a velocity of 667 Hz/s. For the time-change task (Figure 14A), standard sequences were 750 ms in duration (SOA = 250 ms) and thus spanned 500 Hz from start to finish. Comparison sequences were varied around 750 ms (± 5%, 10%, 15%, 20%). For the pitch-change task (Figure 14B), standard glides spanned 1000 Hz, and were thus 1500 ms in duration (SOA = 500 ms). Comparison sequences were varied around 1000 Hz (± 5%, 10%, 15%, 20%).

Stimulus generation and presentation, and response collection were the same as in Experiment 4.

Procedure

The procedure was identical to Experiment 4.

Results

Figure 15 shows proportions of ‘longer’ and ‘more pitch change’ responses as a function of Comparison Level separately for the three comparison velocity conditions for the time-change task (Panel A) and the pitch-change task (Panel B). Similar to Experiment 4, proportions of ‘longer’ and ‘more pitch change’ responses increased with increasing Comparison Level.
(A) Time-change task

**Standard Sequence:** 667 Hz/s

**Comparison Sequence:**
667 Hz/s, 1333 Hz/s, 2000 Hz/s

750 ms ± ΔT

(B) Pitch-change task

**Standard Sequence:** 667 Hz/s

**Comparison Sequence:**
667 Hz/s, 1333 Hz/s, 2000 Hz/s

1500 ms

Figure 14. Stimulus diagram for Experiment 5. Standard sequence velocity was 667 Hz/s. (A) For both the time change task, standard glide duration was 750 ms and standard glide pitch change was 500 Hz. ΔT = ± 5%, 10%, 15%, or 20%, and listeners judged whether the comparison was ‘shorter’ or ‘longer’ than the standard. (B) For the pitch-change task, standard glide duration ΔP = ± 5%, 10%, 15%, or 20%, and listeners judged whether the comparison changed ‘more’ or ‘less’ in pitch than the standard.
Figure 15. Experiment 5. (A) Proportions of ‘longer’ responses and (B) ‘more pitch change’ responses as a function of Comparison Level for the time-change and pitch-change tasks, respectively. There were no systematic effects of Comparison Velocity.
Moreover, there were no systematic effects of Task or Comparison Velocity. The 2 (Task) x 3 (Comparison Velocity) mixed-measures ANOVA confirmed this, revealing only a main effect of Comparison Level, $F(7,238) = 344.89$, $MSE = 0.05$, $p < 0.001$, $\eta^2_p = 0.91$.

Separate 2 (Task) x 3 (Comparison Velocity) mixed-measures ANOVAs for PSE and JND revealed no main effects or interactions ($p_s > 0.15$; $p_s > 0.27$, respectively); see Tables 3 and 4.

**Imputed velocity model**

Values of $w$ obtained by fitting the imputed velocity model to response proportions for the time-change task and pitch-change task were $0.98 \pm 0.01$ and $0.95 \pm 0.02$, respectively. These values differed significantly, $t(34) = 2.20$, $p < 0.05$.

**Summary**

The results of Experiment 5 were consistent with the results of Experiment 4, and indicated that for a standard velocity of 667 Hz/s, listeners’ judgments of duration and pitch change of comparison tone sequences did not depend on Comparison Velocity. Fits of the imputed velocity model to response proportions supported this conclusion; $w$ values were close to 1 for both the time-change and pitch-change tasks. In Experiment 6, standard velocity was 2000 Hz/s. The results were expected to be similar to those obtained in Experiments 4 and 5.

**Experiment 6**

**Method**

**Design**

The design and task were identical to Experiments 4 and 5.

**Participants**

Thirty-four individuals ($n = 27$, female) from a large Midwestern university participated
in exchange for course credit in an introductory psychology class. Listeners self-reported normal hearing and had a range of formal musical training (0 – 14 years, $M = 4.2$ years, $SD = 4.1$ years). Listeners were randomly assigned to complete either the time-change task ($n = 18$) or the pitch-change task ($n = 16$).

Stimuli and apparatus

Stimuli were the same as in Experiments 4 and 5, with the following exceptions. Standard sequences ascended at a velocity of 2000 Hz/s. For the time-change task (Figure 16A), standard sequences were 750 ms in duration (SOA = 250 ms) and spanned 1500 Hz from start to finish. Comparison sequences were varied around 750 ms ($\pm 5\%, 10\%, 15\%, 20\%$). For the pitch-change task (Figure 16B), standard glides spanned 1000 Hz, and were 500 ms in duration (SOA $= 167$ ms). Comparison sequences were varied around 1000 Hz ($\pm 5\%, 10\%, 15\%, 20\%$).

Stimulus generation and presentation, and response collection were the same as in Experiments 4 and 5.

Procedure

The procedure was identical to Experiments 4 and 5.

Results

Figure 17 shows proportions of ‘longer’ and ‘more pitch change’ responses as a function of Comparison Level separately for the three Comparison Velocity conditions for the time-change task (Panel A) and the pitch-change task (Panel B). Similar to Experiments 4 and 5, proportions of ‘longer’ and ‘more pitch change’ responses increased with increasing Comparison Level. Moreover, there were no strong systematic effects of Comparison Velocity on response proportions for either task. Similar to Experiment 4, there were subtle interactions between factors, as indicated by the ANOVA on response proportions.
(A) Time-change task

**Figure 16.** Stimulus diagram for Experiment 3. Standard sequence velocity was 2000 Hz/s. (A) For the time-change task, standard duration was 750 ms and standard pitch change was 1500 Hz. Comparison glides varied around 750 ms ($\Delta T = \pm 5\%, 10\%, 15\%, \text{ or } 20\%$), and listeners judged whether the comparison was ‘shorter’ or ‘longer’ than the standard. (B) For the pitch-change task, standard pitch change was 1000 Hz, and standard duration was 500 ms. Comparison glides varied around 1000 Hz ($\Delta P = \pm 5\%, 10\%, 15\%, \text{ or } 20\%$), and listeners judged whether the comparison changed ‘more’ or ‘less’ in pitch than the standard.
Figure 17. Experiment 6. (A) Proportions of ‘longer’ responses and (B) ‘more pitch change’ responses as a function of Comparison Level for the time-change and pitch-change tasks, respectively. There were no systematic effects of Comparison Velocity.
A 2 (Task) x 3 (Comparison Velocity) x 8 (Comparison Level) mixed-measures ANOVA on response proportions revealed a main effect of Comparison Level, $F(7,224) = 488.00$, $MSE = 0.04$, $p < 0.001$, $\eta^2_p = 0.94$. Proportions of ‘longer’ and ‘more pitch change’ responses increased as a function of Comparison Level. The main effect of Comparison Velocity was also significant, $F(2,64) = 10.76$, $MSE = 0.02$, $p < 0.001$, $\eta^2_p = 0.25$. Proportions of ‘longer’ and ‘more pitch change’ responses were higher for the 667 Hz/s comparison velocity ($M = 0.55 \pm 0.01$) than either the 1333 Hz/s ($M = 0.51 \pm 0.01$), $p < 0.01$, or the 2000 Hz/s comparison velocity ($M = 0.49 \pm 0.01$), $p < 0.001$; response proportions for the latter two velocity conditions did not differ significantly, $p = 0.22$. The Task x Comparison Velocity interaction, $F(2,64) = 5.98$, $MSE = 0.02$, $p < 0.01$, $\eta^2_p = 0.16$, the Task x Comparison Level interaction, $F(7,224) = 4.18$, $MSE = 0.04$, $p < 0.001$, $\eta^2_p = 0.12$, and the Comparison Velocity x Comparison Level interaction, $F(14,448) = 1.81$, $MSE = 0.02$, $p < 0.05$, $\eta^2_p = 0.05$, all reached significance. The Task x Comparison Velocity interaction was driven by a significant difference in mean ‘longer’ versus ‘more pitch change’ response proportions that obtained only for the 1333 Hz/s comparison velocity ($M = 0.54 \pm 0.02$; $M = 0.48 \pm 0.01$, respectively), $t(32) = 2.72$, $p = 0.01$. The Task x Comparison Level interaction was driven by overall steeper slopes for the psychometric functions for ‘more pitch change’ responses relative to ‘longer’ responses. Finally, the Comparison Velocity x Comparison Level interaction was due to an overall leftward shift of the psychometric function for the 667 Hz/s comparison velocity relative to the 1333 Hz/s and 2000 Hz/s comparison velocities that was only apparent in the middle of the Comparison Level range; that is, response proportions converged at the extreme Comparison Level values.

A 2 (Task) x 3 (Comparison Velocity) mixed-measures ANOVA on PSEs supported the analysis for response proportions, revealing a significant main effect of Comparison Velocity,
$F(2,64) = 7.13, MSE = 5.95, p < 0.01, \eta^2_p = 0.18$, and a significant Task x Comparison Velocity interaction, $F(2,64) = 6.15, MSE = 5.95, p < 0.01, \eta^2_p = 0.16$. Overall, PSEs differed between the time-change and pitch-change tasks for the 1333 Hz/s comparison velocity, $t(32) = -2.34, p < 0.05$, and the 2000 Hz/s comparison velocity, $t(32) = 1.76, p = 0.09$, but not for the 667 Hz/s comparison velocity, $t(32) = -0.39, p = 0.70$ (see Table 3).

A parallel analysis on JNDs failed to reveal any significant main effects of interactions ($ps > 0.36$); see Table 4.

**Imputed velocity model**

Values of $w$ obtained by fitting the imputed velocity model to response proportions were $w = 1.00 (\pm 0.001)$ and $w = 0.94 (\pm 0.04)$ for the time judgment and pitch-change tasks, respectively. These values were not significantly different, $t(32) = 1.37, p = 0.18$.

**Comparison of results across standard velocities**

In order to compare results across the three experiments, proportions of ‘longer’ (time-change task) and ‘more pitch change’ (pitch-change task) judgments were entered into a 2 (Task) x 3 (Standard Velocity) x 3 (Comparison Velocity) x 8 (Comparison Level) mixed-measures ANOVA to test for significance of the four-way interaction. The four-way interaction was not significant, $F(28,1372) = 0.95, MSE = 0.02, p = 0.54$, indicating similar patterns of results across Experiments 4, 5, and 6. Indeed, a comparison of Figures 11 – 13 indicates that there were no large, systematic distortions in perceived duration or pitch-change of the sort observed in Experiments 1 – 3. This is consistent with the idea that for discrete tone sequences, velocity information is either less well conveyed or easier to ignore than velocity information conveyed by continuously moving auditory stimuli.

Values of $w$ derived from the imputed velocity model were also examined in a 2 (Task) x
3 (Standard Velocity) between-subjects ANOVA to assess any effects of Standard Velocity. The analysis revealed only a main effect of Task, $F(1,98) = 6.54, MSE = 0.006, p < 0.05, \eta^2_p = 0.06$, which was driven by somewhat higher $w$ values for the time-change task ($w = 0.99 \pm 0.01$) relative to the pitch-change task ($w = 0.95 \pm 0.01$). Neither the main effect of Standard Velocity, $F(2,98) = 0.23, MSE = 0.006, p = 0.80$, nor the interaction, $F(2,98) = 0.43, MSE = 0.006, p = 0.65$, reached significance.

A correlational analysis was not performed for Experiments 4 – 6 because of the restricted range of $w$ values.

Summary

To summarize, the results of Experiment 6 were very similar to the results of Experiments 4 and 5. Although there were subtle differences between average response proportions as a function of Task and Comparison Velocity, these differences were not systematic, and overall the results of Experiments 4 – 6 indicated no systematic over- or underestimation of perceived duration or pitch change of the comparison based on the relationship of the Comparison Velocity to the Standard Velocity.

General Discussion

In Experiments 4 – 6, listeners judged either the total duration or total pitch change of a comparison tone sequence relative to a standard. Within an experiment, the standard velocity was fixed; in Experiment 4, standard velocity was 1333 Hz/s, in Experiment 5 standard velocity was 667 Hz/s, and in Experiment 6, standard velocity was 2000 Hz/s. Comparison velocity varied from trial to trial, taking on one of three values (667 Hz/s, 1333 Hz/s, 2000 Hz/s). In general, listeners accurately judged duration and pitch change regardless of the relative velocities of the standard and comparison sequences. That is, there were no systematic distortions observed
for perceived time or pitch change of the comparison sequence based on the relationship of standard and comparison velocities.

The null results for velocity can be contrasted with the results of Experiments 1 – 3, where large effects of velocity were observed for perceived duration and pitch change. There were two differences between Experiments 1 – 3 and the current experiments that must be considered. First, because of a technical error, the set of velocities used in the current experiments was somewhat faster, and the total duration of standard and comparison sequences was shorter, than in Experiments 1 – 3. The second difference between experiments concerns the continuous versus discrete nature of the auditory stimuli. These differences will be considered in turn.

With respect to velocity and duration differences, a technical error during stimulus generation resulted in the omission of one tone from each sequence; however, the total pitch change of each sequence was left intact. Thus, each tone sequence was 25% shorter in duration than intended, resulting in overall faster tested velocities in Experiments 4 – 6 than in Experiments 1 – 3. In Experiments 1 – 3, the set of velocities was 500 Hz/s, 1000 Hz/s, and 1500 Hz/s, whereas in Experiments 4 – 6, the set of velocities was 667 Hz/s, 1333 Hz/s, and 2000 Hz/s. Although it seems unlikely that the differences in either velocity or duration explain the large qualitative differences in behavioral results between experiments, an additional experiment was conducted to rule out this possibility.

Experiment 4 was replicated with stimulus durations and velocities matched to those of Experiment 1. The velocity of the standard tone sequence was 1000 Hz/s, and the comparison sequence took on velocities of 500 Hz/s, 1000 Hz/s, and 1500 Hz/s. SOAs between tones were identical to Experiment 4, and sequences were composed of five as opposed to four tones. Figure
18 shows the results of the experiment, which included 39 undergraduates from a large Midwestern university (n = 21, time-change task; n = 18, pitch-change task). It is clear from inspection of the figure that there was no effect of Comparison Velocity on response proportions for the time-change task (Panel A). There was a slight increase in proportions of ‘more pitch change’ responses for the 500 Hz/s comparison velocity relative to the 1000 Hz/s and 1500 Hz/s comparison velocities (Panel B). This shift was in the direction predicted by the auditory motion hypothesis; that is, listeners overestimated comparison pitch change when comparison velocity was relatively slow. However, there was no evidence of a shift in response proportions for the relatively fast comparison velocity condition (i.e., 1500 Hz/s). Taking together the results for the time-change and pitch-change tasks, it is considered unlikely that differences in velocity and sequence duration were the source of the large differences observed between experiments reported in Chapters 2 and 3.

The second, and arguably, critical difference between experiments concerns the continuous versus discrete nature of the stimuli. In Experiments 1 – 3, stimuli were tone glides that moved continuously in pitch space, whereas in Experiments 4 – 6, stimuli were discrete tone sequences that only implied motion. Several previous studies provide evidence that continuous and discrete auditory stimuli convey motion information differently. First, Henry and McAuley (in press) studied velocity perception for sounds moving continuously or implying motion in pitch space. They reported that although both time-change and pitch-change information contributed to velocity estimates for continuous tone glides, velocity estimates for discrete tone sequences depended only on time-change information, with no contribution of pitch change. Second, Strybel, Span, and Witty (1998) obtained similar results for discrete sound sequences undergoing apparent motion in auditory physical space. They found that perceived velocity
Figure 18. Results from a replication of Experiment 4; velocities and durations of tone sequences were precisely matched to the glides of Experiment 1. (A) Proportions of ‘longer’ responses and (B) proportions of ‘more pitch change’ responses as a function of Comparison Level for the time change and pitch-change tasks, respectively. There was a slight leftward shift in the psychometric curve for the 500 Hz/s comparison velocity in the pitch-change task. However, overall systematic effects of Comparison Velocity were not observed.
depended on the temporal, but not spatial, properties of the stimulus. Taken together with the results of the experiments presented in Part I of the current dissertation, these studies suggest fundamental differences in the contributions of spatial / pitch and time information to perceived velocity in continuous relative to discrete auditory stimuli. It is important to note that in the current studies, instructions to listeners were to selectively attend to either time or pitch, whereas in the studies of Henry and McAuley (in press) and Strybel, et al. (1998) described above, listeners were instructed to attend to joint pitch-time (i.e., velocity) information. Henry and McAuley (in press) also found that listeners’ were able to separately estimate time-change (i.e., duration) and pitch-change in discrete tone sequences. Thus, discrete auditory stimuli seem to promote selective attention to pitch and time.

These results are at first glance incompatible with previous work by Henry and McAuley (2009) and Henry et al. (2009; see also Cohen, et al., 1954; MacKenzie, 2007; Shigeno, 1986, 1993) on the auditory kappa and tau effects, where pitch-time interdependence was observed in judgments about discrete sequences, despite instructions to attend only to a single dimension. In this regard, it is important to note that in studies of the kappa and tau effects, listeners judged the relative timing or pitch of a single tone embedded in a non-isochronous sequence with irregular pitch intervals. That is, the relative timing and pitch of the target tone were varied on a trial-by-trial basis; thus sequence timing and pitch were not predictable, and sequences were never isochronous (see Figures 1 and 2). One potential explanation is that increasing pitch-time regularity increases the potential for the listener to attend selectively to the pitch and time aspects of the stimulus. Support for this suggestion comes from the studies of Henry (2007) and MacKenzie (2007), who embedded the to-be-judged target tone in a longer six-tone isochronous sequence, thus increasing the availability of regular timing information. Both studies found a
reduced auditory kappa effect, indicating increasing pitch-time independence. In Experiments 4 – 6, both standard and comparison sequences were isochronous and composed of regular pitch intervals on every trial. It is thus likely that strict regularity of the stimuli allowed listeners to effectively ‘split’ attention between the two dimensions. Future research is planned to explore this possibility.

In Part II of this dissertation, the focus is shifted to pitch-time interdependence in attention. Specifically, listeners performed a signal detection task in which performance was enhanced by extrapolating a pitch-time trajectory to the future location of the to-be-detected signal. Stimuli were continuous tone glides or discrete tone sequences that ascended or descended in pitch space. In the following chapters, data will be reviewed supporting the possibility that auditory attention is tuned in pitch and in time, and theories of auditory attention will be reviewed in connection with each of these literatures. Then, some time will be spent addressing a methodological issue that arises when examining perception of expected versus unexpected events; specifically, a relatively large number of observations belonging to an expected condition can lead to an artificial boost in observed perceptual sensitivity. Finally, three experiments will be presented that test the hypothesis that auditory attention is tuned by motion information in auditory stimuli.
PART II
CHAPTER 4: ATTENTIONAL TUNING TO FREQUENCY AND TIME

Previous research has demonstrated that listeners are sensitive to the regularities built into experimental situations, and can exploit those regularities to improve performance. In general, two types of regularities are considered relevant here. First, sensitivity to statistical regularity refers to the tendency of a listener to tune in to more frequently presented events or relationships. For example, in a task that requires detection of a pure-tone signal in noise, listeners enjoy a boost in sensitivity for signals presented at or near the most frequently-presented frequency (Greenberg & Larkin, 1968; Hafter, et al., 1993; Schlauch & Hafter, 1991). Second, structural regularity refers to ordered relationships between frequency intervals and/or time intervals in auditory stimuli that give rise to a predictable frequency-time trajectory. For example, in the time domain, listeners are most accurate to judge events that occur at an expected time based on the rhythmic context within which the event occurs (Barnes & Jones, 2000; Large & Jones, 1999; McAuley & Kidd, 1998). Here, the mechanism by which listeners become more sensitive to a narrow frequency range or a temporal location is referred to as attentional tuning. Of interest is whether attention can be tuned jointly to frequency-time locations by structural regularities in auditory stimuli; the experiments presented here extend work of Crum and Hafter (2008), who showed that listeners are more sensitive to detect a pure-tone signal at a frequency-time location that would be expected based on extrapolation of an ascending tone glide through a segment of noise.

Part II of this dissertation will be organized as follows. First, the remainder of Chapter 4 will be focused on reviewing the literature on attentional tuning. Evidence for attentional tuning to frequency and to time will be presented separately. Then, a case will be made for attentional tuning based on joint frequency-time information. Special attention will be given to the study by
Crum and Hafter (2008), which provides the methodological framework for the studies presented here. Chapter 5 explores a methodological issue that arises when performance is compared for ‘expected’ versus ‘unexpected’ conditions when what is expected is defined as the most frequently presented value of frequency or time, i.e., based on statistical regularity. A single experiment and two simulations are presented. Finally, in Chapter 6, three experiments are reported that explore the main research question addressed in this section, namely whether attention can be tuned on the basis of joint frequency-time information in auditory stimuli.

Attentional Tuning to Frequency

Evidence for attentional tuning to frequency derives its merit from direct comparisons to peripheral masking effects caused by interactions at the level of the basilar membrane. Thus it will be worthwhile to provide a brief introduction to the physiology of hearing as context. First, an overview of peripheral auditory system function will be provided, then studies using the probe-signal method (Greenberg & Larkin, 1968) will be reviewed and compared to peripheral masking studies. Finally, effects of uncertainty on signal detection will be discussed with respect to a two-factor model of signal detection proposed by Johnson and Hafter (1980).

A brief overview of function of the peripheral auditory system

The experience of sound is the result of stimulation of receptors in the cochlea of the inner ear by motion of the basilar membrane (Moore, 2004). Vibration of an object in the environment causes vibration of the surrounding medium (typically air) at a particular frequency. That is, molecules in the air are pushed closer together (compression) and pulled farther apart (rarefaction) in a cyclic fashion; the number of times per second this cycle completes is the frequency of the vibration in Hertz (Hz); the perceptual correlate of frequency is pitch. The
vibrations in the air are transferred to the eardrum (tympanic membrane) and then on to the cochlea by way of the ossicles, three tiny bones in the middle ear.

The cochlea is filled with fluid and divided along its length by two membranes: Reissner’s membrane and the basilar membrane. The motion of the basilar membrane takes the form of a travelling wave, and since the mechanical properties of the basilar membrane vary along its length, different locations along the basilar membrane respond better to vibrations of a particular frequency than others; that is, the travelling wave caused by sounds with different frequencies peaks at different locations along the membrane. Specifically, the basilar membrane is narrow and stiff at the base and wider and less stiff at the apex. Thus, locations near the base respond best to high-frequency vibrations and locations near the apex respond best to low-frequency vibrations.

Receptors (i.e., inner hair cells) are stimulated by a mechanical deformation associated with the movement of the basilar membrane. When the hair cells are deformed, a neural response is stimulated in the connecting nerve that travels to the brain. In the auditory nerve, individual nerve fibers are selectively responsive to frequency; the frequency to which the hair cells show the strongest response is called the characteristic frequency (CF). The receptor also responds to frequencies that are similar to the CF, but sensitivity falls with increasing frequency distance, until the receptor is no longer responsive to vibrations outside of a range referred to as the critical bandwidth (Fletcher, 1940; Scharf, 1961); the critical bandwidth increases with increasing frequency (MacMillan & Schwartz, 1975; Patterson, 1974). The auditory system at the level of the basilar membrane is assumed to act as a bank of bandpass filters, where a filter has two defining characteristics: 1) the center frequency, and 2) the bandwidth.
Converging evidence for the bandpass nature of the peripheral auditory system comes from studies of peripheral auditory masking, which is the phenomenon whereby the threshold for detection of a sound with a particular frequency is increased when another sound is presented concurrently or close in time to the to-be-detected signal (Fletcher, 1940; Harris & Dallos, 1979). For example, when a pure-tone signal is presented concurrently with white noise, the filter centered on the signal frequency responds to the signal, but is also sensitive to the frequencies present in the noise in the range near the signal frequency, i.e., within the critical bandwidth. Thus, the signal-to-noise ratio is effectively decreased, and detection of the pure-tone signal is more difficult. Frequencies falling outside of the critical bandwidth do not affect detectability of the signal (Patterson, 1974; Patterson, Nimmo-Smith, Weber, & Milroy, 1982; Wright & Dai, 1994).

Patterson et al. (1982) used a ‘notched-noise’ technique (Patterson, 1974) to assess auditory filter shape; this technique involves presenting a to-be-detected pure-tone signal concurrently with broadband noise, where the noise is ‘notched’ centered on the signal frequency. That is, frequency components are removed from the broadband noise surrounding the signal frequency. Detection performance is then measured as a function of the width of the notch. The results of Patterson et al. indicated that signal detection performance increased as the range of frequency components removed from the noise was increased, but only until the notch width approximated the width of the critical band; beyond critical bandwidth, increasing notch width did not benefit signal detection. Moreover, consistent with work by Scharf (1961), the width of the critical band was found to increase with increasing frequency.

*The probe-signal method*
Listeners also demonstrate increased detection thresholds for frequencies falling outside of the critical bandwidth in experiments in which peripheral masking is minimal, and auditory filter location and shape are thus assumed to be under attentional control; the term *listening band* is used to refer to auditory filters that are attentional, rather than peripherally determined. In probe-signal experiments (Greenberg & Larkin, 1968; Hafter, et al., 1993; Schlauch & Hafter, 1991), listeners typically detect ‘expected’ signals of a single frequency embedded in noise on the majority of trials, and on a minority of trials, ‘unexpected’ signals of disparate frequencies are presented. Masking by the broadband noise is of a similar degree at all frequencies. However, listeners show diminished ability to detect unexpected relative to expected signals, and detectability decreases with frequency distance from the expected signal. Detection of unexpected signals falls to chance when the frequency distance from the expected signal exceeds approximately the critical bandwidth. Thus, listening band shape derived from probe-signal experiments agrees well with auditory filter shape derived from peripheral masking experiments (Greenberg & Larkin, 1968; Schlauch & Hafter, 1991; 1994); for an exception see Wright and Dai (1994). Figure 19A gives an example of data from a probe-signal experiment.

The probe-signal technique has been useful in clarifying the role of uncertainty in signal detection performance and the ability of different types of cues to alleviate uncertainty. With respect to uncertainty, listeners are capable of monitoring multiple frequency locations for signals, but performance suffers with increasing uncertainty about signal frequency. MacMillan and Schwartz (1975) asked listeners to monitor for two expected signal frequencies simultaneously in a probe-signal task in which the listener did not know which expected signal frequency would be presented on a single trial. Listeners performed as well as in single-frequency detection tasks. However, further increasing uncertainty is detrimental to detection of
Figure 19. (A) Results from a probe-signal experiment, adapted from Greenberg and Larkin (1968). The highest signal detection performance was observed for the expected signal frequency, i.e., 1000 Hz, and performance fell to chance levels when the signal frequency fell outside of the critical bandwidth. (B) A canonical expectancy profile (Barnes & Jones, 2000; Large & Jones, 1999; McAuley & Kidd, 1998). Accuracy for duration judgments is highest when onsets of events marking the to-be-judged intervals (standard or comparison) are expected with respect to a rhythmic context sequence. Accuracy decreases with as event onsets become more unexpected.
the expected signal; Schlauch and Hafter (1991) demonstrated that listeners perform poorly when they must simultaneously monitor four frequency regions relative to when they must monitor only one or two frequency regions.

Two cue types, frequency cues and informational cues, have been shown to alleviate deleterious effects of frequency uncertainty (Howard, et al., 1986). A cue is simply an indicator that, when presented prior to the signal, informs the listener regarding the to-be-monitored frequency location. Frequency cues are defined as auditory cues that precisely match the frequency of the signal (Mondor & Bregman, 1994; Schlauch & Hafter, 1991). Informational cues are informative about the to-be-monitored frequency region because of a learned association of the cue to the signal frequency. For example, an informational cue may be visual; a visual stimulus of one color may signal that the listener should monitor the relatively low frequency region, while a visual stimulus of another color may signal to monitor the higher frequency region (Howard, et al., 1986; Swets & Sewall, 1961). Listeners can also make use of auditory signals presented at a constant frequency ratio to the signal (e.g., a musical fifth) to tune attention in frequency (Hafter, et al., 1993).

Johnson and Hafter (1980) have focused on a distinction between the number / location of listening bands (i.e., center frequencies) that must be monitored and the bandwidth of each of those listening bands. According to their model, listeners are capable of monitoring multiple frequency regions simultaneously or sequentially\(^5\); both frequency cues and informational cues can be used to decide where in frequency the listener should focus on a trial-to-trial basis, that is, center frequency. Moreover, frequency cues presented prior to a trial can effectively narrow the bandwidth of the listening band, thereby reducing the level of the noise passed through the filter, thus increasing the signal-to-noise ratio. Thus, both frequency and informational cues benefit
Auditory Motion for Continuous and Discrete Sounds

Detection performance under conditions of uncertainty; however informational cues are associated with a wider listening band than frequency cues (Hafter, et al., 1993; Howard, et al., 1986; Hübner & Hafter, 1995).

One study of note regarding attentional tuning to frequency was conducted by Mondor and Bregman (1994), who examined the role of attentional tuning to frequency in a duration judgment task. On each trial, listeners were provided with a cue followed by a target tone, the duration of which was to be classified as short or long relative to exemplar durations on which the listeners had received prior training. Three types of cues were presented: 1) valid cues, the frequency of which precisely matched the signal, 2) invalid / close cues, the frequencies of which were close to the signal frequency, and 3) invalid / far cues, the frequencies of which were far from the signal frequency. Duration judgments were faster and more accurate when the cue was valid, and thus provided accurate frequency information; moreover, accuracy declined and reaction times increased as a function of frequency distance of the cue from the signal. Thus, frequency cues presented at the start of each trial acted to narrow the listening band; targets occurring within the to-be-monitored listening band were then more accurately judged with respect to duration. This study highlights an important finding; perception of non-frequency aspects of sounds (e.g., duration) can be influenced by the width of the listening band centered on an expected frequency region.

To summarize, previous research has demonstrated that in the frequency domain, listeners tune into a narrow listening band (Greenberg & Larkin, 1968; Hafter, et al., 1993; Schlauch & Hafter, 1991), the shape of which approximates auditory filters inferred from peripheral masking studies (Patterson, 1974; Patterson, et al., 1982). Attentional tuning to frequency is negatively impacted by uncertainty about the to-be-attended frequency region.
(Hübner, 1993; MacMillan & Schwartz, 1975). However, listeners are capable of using frequency cues and informational cues to counteract the detrimental effects of uncertainty on signal detection (Hafter, et al., 1993; Howard, et al., 1986; Hübner & Hafter, 1995). In particular, frequency cues precisely match the frequency of the to-be-detected target tone, and informational cues indicate the to-be-attended frequency through a learned association between the cue and the signal. One theoretically important study was conducted by Mondor and Bregman (1994), who demonstrated that attentional tuning to frequency affects the accuracy of duration judgments.

The remainder of this chapter reviews the literature on attentional tuning in the time domain and reviews initial evidence that attention can be jointly tuned by frequency-time information.

Attentional Tuning to Time

Early work on attentional tuning to time made use of a temporal probe-signal method. In one study, Egan, Greenberg, and Schulman (1961) varied the duration of the observation interval in which a to-be-detected signal could be presented. The temporal position of the signal was varied from trial to trial, thus longer intervals were associated with increased uncertainty about when the signal would occur. Egan et al. (1961) observed poorer detection performance with longer intervals, indicating that temporal uncertainty negatively impacted signal detection.

Similarly, Leis-Rossio (1986) contrasted signal detection performance for conditions in which signals were always presented at an expected time to conditions in which listeners could not predict the onset time of the signal. Leis-Rossio (1986) observed a benefit in terms of signal detection performance when the signal was presented at an expected point in time.

More recently, evidence for attentional tuning to time has been derived primarily from studies implementing duration judgment tasks where to-be-judged durations are embedded in rhythmic context sequences. Figure 20 shows an example of a time judgment task that involves
(A) Isolated-interval duration judgment task

(B) Embedded-interval duration judgment task

Figure 20. (A) An isolated-interval duration judgment task. The listener judges the duration of the comparison SOA relative to the standard. (B) An embedded-interval duration judgment task. The standard and comparison SOAs are preceded by a rhythmic context sequence. Judgments about the comparison SOA relative to the standard are generally improved when the events marking the standard and comparison intervals occur at expected times with respect to the rhythm of the context sequence.
comparing two isolated durations (Panel A). In isolated-interval tasks, listeners are presented with a standard time interval marked by two auditory events (e.g., pure tones or clicks), followed after a pause by a second comparison time interval (Abel, 1972; Creelman, 1962; Divenyi & Danner, 1977; Getty, 1975). The listener judges the duration of the comparison interval relative to the standard. Figure 20 (Panel B) also shows a time judgment task that involves comparison of two durations, one of which follows a rhythmic context sequence (Panel B). In embedded-interval tasks, an isochronous context sequence precedes presentation of the standard-comparison pair. Critically, in the latter task, the events marking the to-be-judged intervals occur at the expected time relative to the rhythm marked out by the context sequence. The listener must again judge the duration of the comparison interval relative to the standard. In embedded-interval tasks, listeners show reduced temporal discrimination thresholds relative to the isolated-interval task (Halpern & Darwin, 1982; Hirsh, Monahan, Grant, & Singh, 1990).

Notably, listeners show relatively poor discrimination performance when one of the events marking the to-be-judged durations occurs at an unexpected point in time (Barnes & Jones, 2000; Large & Jones, 1999; McAuley & Kidd, 1998). Specifically, ending the standard interval relatively early or late, so that the ending event does not occur at an expected time relative to the context rhythm produces systematic errors in comparison duration judgments (Barnes & Jones, 2000; Large & Jones, 1999); beginning the comparison interval relatively early or late produces a similar pattern of results (McAuley & Jones, 2003; McAuley & Kidd, 1998). This systematic pattern of sensitivity change as a function of temporal distance from an on-time position defined by the context rhythm is referred to as an expectancy profile (Figure 19B). It is perhaps useful to draw an analogy between the expectancy profile and the shape of the frequency listening band derived from probe-signal tasks (Figure 19A). Performance is best when a to-be-
judged event occurs \textit{where} in frequency or \textit{when} in time it is expected to, and sensitivity falls off as the critical event strays from its expected frequency / time.

Jones and colleagues (Barnes & Jones, 2000; Jones, 1976, 2004; Large & Jones, 1999; McAuley, 1995; McAuley & Jones, 2003) have offered a theory of rhythmic attention that provides an account of attentional tuning to time with respect to both temporal location and tuning precision (analogous to center frequency and bandwidth, respectively). This theory of dynamic attending is biologically based, and attributes to attention many qualities shared with \textit{circadian rhythms}, an example of which is the 24-hour sleep-wake cycle that is coupled to the environmental light-dark cycle (Pittendrigh & Daan, 1976). Dynamic attending theory is built on three key assumptions: 1) \textit{Driving rhythms}: Environmental events take place in time and can be defined according to their temporal structure; that is, events occur at a particular rate and have varying degrees of temporal regularity. 2) \textit{Driven rhythms}: Humans are inherently rhythmic organisms, exhibiting endogenous neural periodicities ranging in terms of rate from milliseconds to years. 3) \textit{Entrainment}: Driven rhythms of the organism can become coupled, or synchronized, with driving rhythms of the environment through adjustments in the relative phase and period of internal rhythms; entrainment extends to attentional rhythms (Jones, 1976, 2004). The treatment of dynamic attending theory given here is conceptual in nature. The reader is referred to McAuley (1995) and Large and Jones (1999) for separate mathematical treatments of the theory; see also Jones (2004).

A key proposition of dynamic attending theory is that rather than being uniformly distributed in time, attentional energy waxes and wanes, with ‘peaks’ (analogous to listening band center) coinciding with the critical portions of the signal (i.e., event onsets). Over time, as synchronization improves, attentional peaks become more precisely aligned with events onsets,
and the width of attentional pulses narrows. Two aspects of attending distinguished by dynamic attending theory promote entrainment of attention to environmental rhythms. Anticipatory attending refers to shifts in attentional timing such that peaks in attentional energy are expected to coincide with future event onsets; anticipatory attending is promoted when auditory patterns are characterized by a high degree of temporal regularity. Reactive attending refers to sudden adjustments of attentional timing that occur in response to unexpected event onsets (Jones & Boltz, 1989).

Better discrimination performance for embedded-interval versus isolated-interval duration judgments is explainable in terms of entrainment of attention, i.e., attentional tuning, by the regular rhythm of the context sequence (McAuley, 1995). Alignment of attentional peaks with critical events affords precision in judgments about the timing of future events. Events occurring at unexpected early or late time points (i.e., standard endings or comparison beginnings) fail to coincide with attentional peaks, and moreover trigger sudden adjustments, i.e., reactive attending. Thus, the attentional rhythm is disrupted and the width of the attentional pulse increases in response to unexpected event onsets, decreasing the accuracy with which upcoming events are perceived (Barnes & Jones, 2000; Large & Jones, 1999; McAuley & Kidd, 1998). Temporal irregularity in context sequences also decreases discrimination performance (Barnes & Jones, 2000; Drake & Botte, 1993; McAuley & Jones, 2003; Schulze, 1978), as irregular rhythms do not allow for efficient entrainment.

One notable study that parallels that of Mondor and Bregman (1994) was conducted by Jones and colleagues (2002); recall that Mondor and Bregman demonstrated that judgments about duration were more accurate when listeners were provided with a frequency cue that tuned attention to the frequency location of the to-be-judged event. In the study by Jones et al.,
listeners heard 10-tone patterns and judged whether the pitch of the last tone (comparison) was *lower, same, or higher* than the first tone (standard), ignoring isochronous interpolated tones which were assigned random pitches. When the comparison tone occurred on time relative to the rhythm instantiated by the interpolated tones, pitch judgment performance exceeded that observed when the comparison tone occurred early or late, resulting in an expectancy profile. Thus nontemporal (i.e., pitch) judgments were shown to be influenced by the degree of synchrony of an attentional peak with the to-be-judged event. That is, the presence of the interpolated tones acted as a cue to *when* in time the to-be-attended event would occur, and strict isochrony encouraged fine temporal tuning of the attentional pulse. When irregularity was introduced into the pattern by randomizing the temporal intervals marked out by the interpolated tones, there was no effect of the temporal position of the comparison tone on pitch judgments (cf. Barnes & Jones, 2000).

To summarize, previous research has demonstrated that in the time domain, listeners tune into expected times of event onsets (Large & Jones, 1999; McAuley, 1995); judgments about duration are more accurate when events marking to-be-judged time intervals occur at expected times based on a rhythmic context sequence. Attentional tuning to time is negatively impacted by temporal irregularity, i.e., uncertainty about the to-be-attended temporal position (Barnes & Jones, 2000; Drake & Botte, 1993; McAuley & Jones, 2003; Schulze, 1978). Jones and colleagues have put forward a theory of rhythmic attention that accounts well for temporal attentional tuning results (Jones, 2004; Large & Jones, 1999); dynamic attending theory assumes that attentional rhythms become coupled to external stimulus rhythms through entrainment and thereby facilitate attending to a future expected time point. A study parallel to that conducted by Mondor and Bregman (1994) demonstrated that attentional tuning to time affects the accuracy of
pitch judgments (Jones, et al., 2002). Next, evidence will be reviewed that supports the possibility that attention can be jointly tuned by frequency-time information; special attention will be given to a study conducted by Crum and Hafter (2008), which provides the methodological basis for the experiments reported in Chapter 6.

**Attentional Tuning to Frequency and Time**

In the preceding sections, two studies were discussed that demonstrated on the one hand that attentional tuning to frequency can influence time perception (Mondor & Bregman, 1994), and on the other hand that attentional tuning to time can influence pitch perception (Jones, et al., 2002). A fundamental research question that is hinted at but not explicitly addressed by these studies is: to what extent do frequency and time structures in time-varying stimuli interact to jointly tune attention? This is the question that is addressed by Part II of this dissertation.

The auditory motion hypothesis predicts joint attentional tuning by frequency and time structure in auditory stimuli, and in particular, suggests that listeners should be capable of using frequency-time trajectory information to tune attention to a future event location. Studies from vision indicate that observers readily make use of visual trajectory information to predict the future location of an occluded object (Churchland, Chou, & Lisberger, 2003; Orban de Xivry, Bennett, Lefevre, & Barnes, 2006; Yantis, 1995). Moreover, in audition, listeners have been shown to be capable of extrapolating a continuous or discrete auditory stimulus through an interrupting noise burst (Ciocca & Bregman, 1987; Kluender & Jenison, 1992; Pollack, 1977). Most relevant here, a recent study by Crum and Hafter (2008) provides a compelling demonstration of the propensity of listeners to tune attention to an expected frequency-time location on the basis of a trajectory-based cue. This study is the foundation for the experiments to be presented next, and so will be reviewed in detail.
Crum and Hafter (2008) demonstrated attentional tuning based on extrapolation of a frequency-time trajectory using a modified probe-signal method (Greenberg & Larkin, 1968). Listeners heard a cue that consisted of an ascending linear tone glide followed by a noise burst that varied in duration (50 ms – 200 ms); the linear velocity of the cue was fixed at 3000 Hz/s. On 50% of trials (signal trials), the noise burst was followed by a to-be-detected 100-ms pure tone signal, and on 50% of trials (noise trials), the signal was absent (see Figure 29). Stimuli, including the cue and signals, were presented against a continuous background of low-level broadband noise. The ‘expected’ signal frequency was determined by extrapolation of the frequency-time trajectory of the glide through the variable-duration noise burst; unexpected signal frequencies were varied symmetrically around the location of the expected frequency, taking on values of ± 10% or 20% of the expected signal frequency. The expected signal frequency was presented on 60% of signal trials, and each of the four unexpected frequency values was presented on 10% of signal trials. Starting frequency and duration of the cue were varied randomly from trial to trial; starting frequency was sampled from a uniform distribution ranging between 300 Hz and 1200 Hz, and duration took on values ranging between 300 ms and 600 ms. Randomizing both starting frequency and glide duration ensured that listeners could not adopt a strategy of listening to a narrow frequency region to accomplish the task. Listeners indicated whether a signal was present following the noise burst by responding ‘yes’ (signal present) or ‘no’ (signal absent).

A summary of results from Crum and Hafter (2008) is reproduced in Figure 21. Overall, listeners demonstrated heightened sensitivity to detect the expected signal frequency based on extrapolation of the cue through the noise burst, and relatively poor sensitivity for unexpected frequencies; detection fell off as the frequency distance from the expected signal increased. The
Figure 21. Reproduced results from Crum and Hafter (2008). Perceptual sensitivity ($d'$) is shown as a function of Signal Frequency separately for each Noise Burst Duration range. Overall, sensitivity to detect the signal was highest when the signal occurred at the expected frequency based on extrapolation of the tone glide cue through the noise burst; sensitivity decreased with increasing distance from the expected signal frequency. For the longest noise burst (150-200ms), sensitivity was highest for signal frequencies somewhat lower than the expected frequency location.

detection data can be compared to behavioral data from the standard probe-signal method (Figure 19A). It should be noted that starting frequency of the cue, cue duration (Experiment 2), and noise burst duration were roved from trial to trial. Thus, targeted listening in the region of the expected signal frequency could not have been accomplished by any strategy other than extrapolating the frequency-time trajectory of the cue through the noise burst. One question that is raised by these results in the context of the experiments reported in Chapters 2 and 3 of this
dissertation is whether listeners would similarly extrapolate the frequency-time trajectory implied by a discrete tone sequence cue through a variable-duration noise burst. This question will be addressed by the experiments reported in Chapter 6.

An additional finding of interest was that longer duration noise bursts were associated with a shift in the center of the listening band to lower frequencies. That is, listeners better detected signals whose frequencies were lower than expected based on extrapolation of the cue through the noise burst when the noise burst was relatively long. A follow-up experiment (Crum & Hafter, Experiment 3) suggested that this ‘undershoot’ of the cue trajectory for long noise bursts was driven by a shift in the center of the listening band rather than by an increase in listening band width. The authors proposed a mechanism for the observed underestimation of the expected frequency when noise burst duration was relatively long. Specifically, it was suggested that listeners average current and past velocity information contained within a sliding temporal window, estimated in their study to have duration equal to 365 ms. When the cue terminates at the onset of the noise burst, the velocity estimate then quickly converges to zero, causing an underestimation in the predicted signal frequency based on velocity information. On this view, extrapolation effectively loses momentum upon termination of the cue, and the extent to which expected signal frequency is underestimated varies with extrapolation duration. A specific prediction was made by Crum and Hafter (2008) with respect to cues with descending frequency-time trajectories. If, upon termination of the cue, estimated velocity converges to zero, the consequences for descending cues should mirror the results for ascending cues. Specifically, Crum and Hafter (2008) predicted that the center frequency of the auditory filter should be higher than the frequency location of the expected signal, and listeners should be better to detect
higher-frequency signals following long noise bursts when cues are descending. That is, listeners should overestimate the expected signal frequency following a descending cue.

Previous work by Henry and McAuley (2009) and Henry et al. (2009) suggests a different outcome for detection of signals following descending cues. The authors have suggested that listeners make use of an auditory gravity-like principle in auditory stimuli that move or imply motion in pitch space. Specifically, the auditory kappa and tau effects were compared for ascending and descending tone sequences, and for both effects, interactions between sequence velocity and direction (ascending versus descending) were observed. For the kappa effect, where the task involved judging the timing of a target tone in a three-tone sequence, the magnitude of perceptual distortion was seen to be larger for descending relative to ascending sequences, specifically for the slowest velocity condition (8ST / 1600 ms). Similarly, for the tau effect, where the task involved judging the relative pitch of a target tone embedded in three-tone sequences, the overall magnitude of the tau effect was magnified for descending relative to ascending sequences, and this was especially true for the two slowest velocity conditions (16 ST / 1000 ms, 16 ST / 1600 ms). It was suggested that descending sequences were heard as accelerating, based on an auditory gravity hypothesis, and thus velocities on the fringe of the auditory motion range were pushed into the range within which listeners picked up on trajectory information.

Consistent with this view, Hubbard (1995b, 2005) has suggested that knowledge of invariant physical forces affects displacement of visual stimuli in memory. In visual representational momentum tasks, observers view a sequence of objects implying motion and must indicate whether a subsequent probe object was presented in the same or different location as the final sequence element. Observers tend to respond ‘same’ more often when the probe
object is displaced farther along the implied motion trajectory, reflecting extrapolation of object motion implied by the sequence. Hubbard and others have shown gravity-like effects (Hubbard, 1990, 1995b; Hubbard & Bharucha, 1988; Shiffrar & Shepard, 1991), in that the degree of displacement is larger for sequences implying downward as opposed to upward motion in the vertical plane. Moreover, for object sequences moving in the horizontal plane, memory for the position of the final sequence element is displaced downward, consistent with the operation of an internalized gravity representation on perception of and memory for an object implying motion. Hubbard (1995a) provided tentative evidence that a gravity-like representation may also act on memory for the frequency location of auditory stimuli in a similar task that involved indicating whether a probe tone was the same or different pitch as the final tone of an ascending or descending four-tone sequence. Memory for the final tone was displaced in the direction of implied motion, as indicated by increased proportions of same responses when the final tone was higher or lower than the penultimate tone, for ascending and descending sequences, respectively. Hubbard (1995a) reported greater auditory representational momentum for descending sequences than for ascending sequences.

In a more musical domain, Larson and VanHandel (2005) proposed that during music listening, a ‘gravitational force’ that is grounded in knowledge of gravity in the physical world acts on expectations of listeners. They tested listeners’ judgments of completion for tone triads, in which the final tone either confirmed or violated expectations based on musical gravity, magnetism, and inertia. Listeners judged triads as more complete when the triad ended with a descending pitch interval, consistent with an auditory gravity hypothesis. Thus, an additional question raised by these results concerns extrapolation of ascending versus descending cues to
tune attention to an expected frequency-time location. This question will be addressed in Chapter 6.

Chapter 5 addresses an important methodological issue that arises when comparing sensitivity to detect signals that are ‘expected’ versus ‘unexpected’ when the numbers of expected and unexpected signals differ. Indeed, assessing effects of expectation, almost by definition, requires presenting more expected than unexpected stimuli; however, the experiment and simulations reported next indicate that differences in observed $d'$ may be present that do not reflect differences in true sensitivity of the observer.
CHAPTER 5: ASSESSING EFFECTS OF EXPECTATION ON PERCEPTUAL SENSITIVITY

Within the cognitive psychology literature, there have been a variety of studies that have reported an advantage for stimuli that are expected versus stimuli that are unexpected. A notable example is the seminal work of Posner and colleagues showing that observers respond more quickly to a brief visual stimulus presented in an expected spatial location than to the same stimulus at an unexpected spatial location (Posner, 1980; Posner, et al., 1980). In the auditory psychophysics literature, a similar logic underlies research on frequency tuning using the probe-signal method, where listeners are more sensitive to detect a tone whose frequency is presented frequently, and is therefore expected, relative to tones whose frequencies are infrequently presented, and are therefore unexpected (Greenberg & Larkin, 1968; Hafter, et al., 1993; Schlauch & Hafter, 1991). In this work, it is critical that expected frequency is defined as such because it is the signal frequency presented on the majority of trials. However, the results of Experiment 7, reported below, suggest that extra care should be taken when interpreting advantages conferred to expected versus unexpected stimuli when expectation is manipulated by presenting more expected than unexpected stimuli. Specifically, it is possible to observe heightened sensitivity to detect an expected signal relative to an unexpected signal when there are no differences in true sensitivity per se. Based on the results of Experiment 7, two simulations were conducted to determine the factors that contributed to this potential artifact, and to determine the appropriate parameters for the studies reported in Chapter 6, which used a modified probe-signal technique.

Experiment 7

Overview

Listeners detected pure-tone signals in white noise; signals randomly took on 1 of 45
frequencies. The signal detection metric $d'$ quantified sensitivity to detect the pure-tone signals. For each listener, two analyses were conducted. First, an unbalanced analysis was conducted, which was modeled after the study by Crum and Hafter (2008), where 60% of signal trials were expected and 10% of signal trials belonged to each of four unexpected conditions. Second, a balanced analysis was conducted in which equal numbers of signal trials were assigned to each of two samples to provide an estimate of the degree to which estimated $d'$ values could be expected to differ by chance. To preview, significant differences were observed in estimated values of $d'$ between samples for the unbalanced, but not the balanced analysis.

**Method**

**Design**

The design was a 2 (Analysis Type: unbalanced, balanced) x 2 (Trial Type: signal, noise) within-subjects factorial. Listeners detected a pure-tone signal in noise by responding ‘yes’ (signal present) or ‘no’ (signal absent); the signal was present on 50% of trials. Two analyses were conducted on each listener’s data set; first an unbalanced analysis assigned 60% and 10% of signal trials to two samples, sample A and sample B, respectively. A second, balanced analysis assigned 20% of signal trials to both sample A and sample B. For both analyses, $d'$ was estimated for each sample as described below.

**Participants**

Twelve undergraduate psychology students ($n = 8$, female) from a large Midwestern university participated in exchange for course credit; all were naïve to the purpose of the experiment. Listeners self reported normal hearing and had a range of formal musical training (0 – 10 years, $M = 5.8$, $SD = 3.1$).

**Stimuli and apparatus**
A schematic of stimuli for Experiment 7 is shown in Figure 22. A 1000-ms segment of 
~60-dB white noise was followed by relatively loud, ~70-dB white-noise burst of variable 
duration (175 ms, 250 ms, 325 ms). A 100-ms segment of ~60-dB white noise followed the noise 
burst. On 50% of trials, a quiet (~40 dB) 100-ms pure-tone signal was embedded in the final 
segment of noise. The frequency of the signal took on one of 45 values, ranging between 850 Hz 
and 2150 Hz.

Stimuli were generated off-line using MATLAB software (The Mathworks, Inc.). 
Stimulus presentation and response collection were controlled using E-Prime 2.0.8.73 software 
(Psychology Software Tools, Inc.) running on a Dell Optiplex computer. Auditory stimuli were 
presented over Sennheiser HD 280 Pro headphones (Old Lyme, CT), and responses were made 
using a serial response box.

Procedure

Participants heard recorded instructions, then completed a short 18-trial training block 
with corrective feedback followed by three experimental blocks with no feedback. On each trial, 
participants heard a single stimulus, responding ‘yes’ if they detected a pure-tone signal in the 
interval following the variable-duration noise burst and ‘no’ if they did not. Signals were 
presented on 50% of all trials (i.e., signal trials). In each block, participants responded to 60 trials 
(20 per noise burst duration). Overall, participants completed a total of 360 trials in 6 
experimental blocks, responding 4 times to each signal frequency. Participants were permitted to 
take short rest breaks between experimental blocks. The experiment lasted approximately 45 
minutes.

Data analysis

Sensitivity to detect the signals was quantified by the signal detection metric \(d'\), which
Figure 22. Stimulus and task diagram for Experiments 7 and 8. Listeners detected a pure-tone signal following a variable-duration noise burst by responding ‘yes’ (signal present) or ‘no’ (signal absent). The signal was present on 50% of trials (signal trials). The signal, noise burst, and 1000 ms preceding the noise burst was presented against a background of low-level broadband noise to mask the signal.
was calculated according to the formula $d' = z(\text{Hit Rate}) - z(\text{False Alarm Rate})$. The hit rate was defined as the proportion of times a listener responded ‘yes’ when a signal was present (i.e., signal trials), and the false alarm rate was defined as the proportion of times a listener responded ‘yes’ when a signal was absent (i.e., noise trials). When hit rates or false alarms rates were equal to 0 or 1, a correction was applied before the value was z-transformed that replaced proportions of 0 and 1 with $1/2N$ and $1-(1/2N)$, respectively, where N corresponds to the number of signal or noise trials presented for that condition (MacMillan & Creelman, 2005).

In order to examine potential biases in the estimates of $d'$ associated with samples with differing numbers of observations, two analyses were conducted for each listener. For both analyses, only one third of the total trials were considered; the reason for this was to match the numbers of observations in each sample to Experiments 9 and 10, where the total number of trials was divided between three noise burst duration conditions. First, for the unbalanced analysis, 60% and 10% of each listener’s signal trials were randomly assigned to two to-be-compared samples, referred to as sample A and sample B, respectively. For the current experiment, this meant that the number of sample A observations, $n_A$, was 36, and the number of sample B observations, $n_B$, was 6. Second, for the balanced analysis, 20% and 20% of each listener’s signal trials were randomly assigned to two samples; thus $n_A = n_B = 15$. No trials were assigned to both sample A and sample B. Separate hit rates were calculated for each sample, but only a single false alarm rate was calculated across all noise trials (Werner, Parrish, & Holmer, 2009); the number of noise trials used in both analyses was 60.

Results and Discussion

Figure 23 shows $d'$ for each sample according the unbalanced and balanced analyses. It is clear from the figure that values of $d'$ were similar across samples for the balanced analysis.
Figure 23. Experiment 7. Observed $d'$ values for sample A (white bars) and sample B (shaded bars) are shown for the balanced and unbalanced analyses. No differences in sensitivity were observed between samples for the balanced analysis, but a significant difference was observed for the unbalanced analysis, with $d'$ for sample A being larger than for sample B.

However, for the unbalanced analysis, estimated values of $d'$ were larger for the sample with more trials (60%, sample A) than for the sample with fewer trials (10%, sample B). Separate paired-samples $t$-tests were conducted to compare the two samples in each analysis type. Estimated $d'$ did not differ between samples for the balanced analysis ($d'_A = 2.89 \pm 0.15$; $d'_B = 2.95 \pm 0.013$), $t(11) = -1.48$, $p = 0.17$; however, this difference was highly significant for the unbalanced analysis, $t(11) = 12.81$, $p < 0.001$; observed $d'$ was larger for sample A ($d' = 3.34 \pm 0.19$), which included a larger number of observations, than for sample B ($d' = 2.62 \pm 0.17$).
Thus, the results of Experiment 7 indicated that the signal detection measure of sensitivity, $d'$, was influenced by unequal numbers of observations in two to-be-compared samples, leading to differences in estimated sensitivity that do not reflect differences in true sensitivity.

In order to isolate the experimental factors contributing the magnitude of the artifact, and to pin down methodological details for subsequent experiments that avoid this artifact, two simulations were conducted. The first simulated performance of a hypothetical observer according to a signal detection model. True sensitivity was varied parametrically and was the same for conditions with more and less trials. As in the unbalanced analysis of Experiment 7, the ratio of trials in the two conditions was 6:1, but the absolute number of trial presentations was varied. Two corrections for proportions of 0 and 1 were tested. The second simulation estimated the difference between values of observed sensitivity for conditions with more and less observations (sample A and sample B, respectively), where the number of observations in sample B, $n_B$, varied between 1 and $n_A$, the number of observations in sample A.

**Simulation 1**

The aim of this simulation was to investigate the factors that potentially introduce bias into estimates of $d'$. A second, related aim, was to determine the levels of each factor that, in combination, eliminate the possibility for an artifactual difference in observed sensitivity for conditions with more (sample A) and fewer observations (sample B). There were three variables of interest: 1) true sensitivity ($d'$) of the observer to detect a pure-tone signal in noise, 2) the absolute number of observations, holding the ratio of sample A to sample B trials, $n_A:n_B$, constant at 6:1, and 3) the correction applied to proportions of 0 and 1 prior to applying a z-transformation.

Miller (1996) previously found that the magnitude of bias of estimated $d'$, i.e., the degree
of discrepancy between true and observed $d'$, depended on true $d'$ and the number of observations, while the direction of the bias depended on the correction applied to proportions of 0 and 1. Specifically, larger values of true $d'$ led to a larger bias, as did smaller absolute numbers of observations. Moreover, Miller (1996) examined three corrections for proportions of 0 and 1; replacing values of 0 and 1 with 0.05 and 0.95 (Murdock Jr. & Ogilvie, 1968), respectively, or omitting cells with proportions of 0 and 1 led to an underestimation of true $d'$. A third correction that involved replacing values of 0 and 1 with 0.0001 and 0.9999 led to a large overestimation of true $d'$. Particularly relevant for the current work, Miller (1996) also indicated that comparing across conditions with unequal numbers of observations lead to significant differences in observed $d'$ that did not reflect differences between values of true $d'$. In his simulation, the ratio of sample A to sample B trials, $n_A:n_B$, was 4:1, and Miller (1996) observed significant differences in observed $d'$ between sample A and sample B on 93% of iterations. The current simulation builds on the report of Miller (1996) by examining two corrections that were not examined by Miller and assessing differences in observed $d'$ without differences in true $d'$ when $n_A:n_B$ was 6:1.

In the current simulation, true sensitivity took on eight values: $d' = 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0$. The ratio of sample A to sample B trials, $n_A:n_B$, was fixed at 6:1, but the absolute number of trial presentations was simulated at five values: 6:1, 60:10, 120:20, 240:40, 480:80. Finally, two corrections were applied to proportions of 0 and 1. The first involved replacing proportions of 0 and 1 with $1/2N$ and $1-(1/2N)$, respectively, where $N$ refers to the total number of trials corresponding to the cell being replaced (Macmillan & Creelman, 2005). The second correction involved adding 0.5 to raw numbers of hits and false alarms and 1 to the total number of signal and noise trials before calculating proportions of hits and false alarms (Hautus, 1995).
The latter method has previously been shown to reduce bias in estimates of sensitivity (Brown & White, 2005).

The procedure involved simulating the behavior of a listener according to a signal detection model assuming unbiased responding. On each trial, a stimulus was drawn from one of two normal distributions corresponding to the signal-plus-noise and noise distributions; the distributions were assumed to have equal variance, \( \sigma = 1 \). The mean of the signal-plus-noise distribution was \( 1/2 \cdot d' \), and the mean of the noise distribution was \( -1/2 \cdot d' \), where \( d' \) was varied parametrically. Signal trials for sample A and sample B were drawn from the same distribution, i.e., the signal-plus-noise distribution. If the value of the sampled stimulus exceeded the response criterion \( (c = 0) \), the listener responded ‘yes’ (signal present), and if the value of the stimulus fell below the response criterion, the listener responded ‘no’ (signal absent). Hits were defined as ‘yes’ responses to signal trials, and false alarms were defined as ‘yes’ responses to noise trials. Hit rates were calculated separately for the signal trials in sample A and sample B, but a single false alarm rate was calculated across all noise trials that was shared by the samples (Werner, et al., 2009). The number of noise trials was always equal to \( n_A + 4 \cdot n_B \). For each value of true \( d' \), for each absolute number of presentations, the simulation was performed for 100 iterations; reported \( d' \) values are averaged over iterations.

Results and Discussion

Figures 24 and 25 show estimated \( d' \) values for sample A (shaded bars) and sample B (white bars) signal conditions as a function of true \( d' \). Separate plots (Panels A-E) are shown for the five levels of number of trial presentations, \( n_A:n_B = 6:1, 60:10, 120:20, 240:40, 480:80 \). Figure 24 shows results when proportions of 0 and 1 are replaced using the correction suggested by MacMillan and Creelman (2005), and Figure 25 shows results for the correction suggested by
Figure 24. Simulation 1. Observed $d'$ values are shown for sample A (shaded bars) and sample B (white bars) as a function of true $d'$. Separate plots are shown for each number of trial presentations. Proportions of 0 and 1 were corrected using the method of MacMillan and Creelman (2005). Asterisks mark significant differences between observed sensitivity for sample A and sample B.
Figure 25. Simulation 1. Observed $d'$ values are shown for sample A (shaded bars) and sample B (white bars) as a function of true $d'$. Separate plots are shown for each number of trial presentations. Proportions of 0 and 1 were corrected using the method of Hautus (1995). Asterisks mark significant differences between observed sensitivity for sample A and sample B.
Hautus (1995). Of interest was the difference between values of estimated $d'$ for sample A and sample B for each value of true $d'$ and each number of trial presentations. In both figures, pairs of bars are marked with an asterisk (*) to indicate a significant difference between estimated $d'$ values for sample A and sample B in the absence of differences in true $d'$. Family-wise $\alpha$-levels were adjusted for multiple comparisons according to a Bonferroni correction; for each combination of number of trial presentations and correction, family-wise $\alpha$ was set to 0.10 (per-comparison $\alpha = 0.0125$).

Overall, Simulation 1 yielded four main results. First, with very low numbers of observations ($n_A = 6$, $n_B = 1$), observed $d'$ underestimated true $d'$ for both sample A and sample B, particularly when values of true $d'$ were large; this was true for both corrections. Second, the discrepancy between observed $d'$ values for sample A and sample B increased with increasing values of true $d'$. Third, increasing the number of observations decreased the discrepancy between sample A and sample B, especially when true $d'$ was relatively small. Finally, the corrections suggested by MacMillan and Creelman (2005) and Hautus (1995) performed similarly in terms of the number of true $d' \times$ number of presentations combinations for which estimated $d'$ differed between sample A and sample B. It is notable that the correction of Hautus (1995) has previously been suggested to more accurately approximate true $d'$ relative to the MacMillan and Creelman correction (Brown & White, 2005). However, the current results suggest that, at least when the number of presentations differs between conditions, the MacMillan and Creelman (2005) and Hautus corrections perform similarly. Taken together, the results suggest that the artificial difference between samples in terms of sensitivity can be reduced by presenting a large number of trials in each condition (e.g., $n_A:n_B = 480:80$) and calibrating the task such that baseline sensitivity is relatively low. To further explore this issue, a
second simulation was conducted that provides a more analytical treatment of the problem of bias in estimation of $d'$ for conditions with different numbers of observations.

**Simulation 2**

The aim of the second simulation was to determine analytically the number of observations required for sample B to eliminate the difference between observed $d'$ values for sample A and sample B given equivalent values of true $d'$. The number of sample A signal trials, $n_A$, and true $d'$ were varied parametrically and took on the same values tested in Simulation 1 ($n_A = 6, 60, 120, 240, 480$; true $d' = 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0$). For each value of $n_A$, $n_B$ took on every value between 1 and $n_A$; for each $n_A-n_B$ pair, the number of noise trials was equal to $n_A + 4* n_B$. For each combination of $n_A$ and $n_B$, the probabilities of observing hit rates equal to $h/ n_A$ and $h/ n_B$ were calculated according to the binomial distribution, where $h$ varied between 1 and $n_A$ and 1 and $n_B$, respectively. Similarly, the probability of observing a false alarm rate equal to $f/(n_A + 4* n_B)$ was calculated in the same way, where $f$ varied between 1 and $n_A + 4* n_B$.

Consider for example a true $d' = 2.0$, and assume zero bias ($c = 0$). In this case, only one hit rate / false alarm rate pair is possible: $HR = 0.84$, $FAR = 0.16$. The probability of observing $h$ hits is given by the formula:

$$ p(HR) = \binom{s}{h} p^h (1 - p)^{s-h} $$

(3)

where $\binom{s}{h} = \frac{s!}{h!(s-h)!}$. Here, $s$ refers to the number of signal trials (either $n_A$ or $n_B$), $h$ refers to the number of hits, and $p(HR)$ corresponds to the probability of observing hit rate equal to $h/s$ given true $d'$. Thus for the $n_A$: $n_B$ pair 6:1, the probability of observing 3 hits in $n_A = 6$ trials is given by

$$ p(HR) = \binom{6}{3} \cdot 0.84^3 (1 - .84)^{6-3} = .05. $$

This was repeated for all values of $h$ between 1 and $n_A$. Similarly, the probability of obtaining $h$
hits for every value of \( h \) from 1 to \( n_A \) was determined for sample B according to the same formula. Next, for each combination of \( n_A \) and \( n_B \), the probability of obtaining \( f \) false alarms was calculated given true \( d' \) according to the formula

\[
p(FAR) = \binom{i}{f} p^f (1 - p)^{i-f}
\]

where \( i \) refers to the number of noise trials \( (n_A + 4*n_B) \), \( f \) refers to the number of false alarms, and \( p(FAR) \) corresponds to the probability of observing a false alarm rate equal to \( f/i \) rate given true \( d' \). For \( n_A: n_B = 6:1 \), the total number of noise trials was 10. Given \( FAR = 0.16 \), the probability of observing 3 false alarms was

\[
p(FAR) = \binom{10}{3} .16^3 (1 - .16)^{10-3} = .15.
\]

For \( A:B = 6:1 \), observing 3 hits for sample A trials and 3 false alarms yields an observed \( d' = z(0.5) + z(0.3) = 0.52 \) for sample A; the joint probability \( p(HR)*p(FAR) = .0075 \) means that there is a .75% chance of observing \( d' = 0.52 \) for sample A given \( d' = 2 \) and \( n_A: n_B = 6:1 \).

For each combination of \( n_A \) and \( n_B \), values of observed \( d' \) were calculated for every possible raw value of hits and false alarms from 1 to \( n_A \) for sample A, 1 to \( n_B \) for sample B, and 1 to \( n_A + 4*n_B \) for noise trials. The result was an \( n_A \times (n_A + 4*n_B) \) matrix of \( d' \) values for sample A, and an \( n_B \times (n_A + 4*n_B) \) matrix of \( d' \) values for sample B. Moreover, corresponding probability matrices were computed according to the procedure described above, yielding an \( n_A \times (n_A + 4*n_B) \) matrix of probabilities of observing a particular \( d' \) value for sample A, and a \( n_B \times (n_A + 4*n_B) \) matrix of probabilities of observing a particular \( d' \) for sample B. Separately for sample A and sample B, the \( d' \) and probability matrices were multiplied element-by-element then summed to yield the most likely values of observed \( d' \) for the two samples given true \( d' \). Observed \( d' \) for sample B was then subtracted from observed \( d' \) for sample A to yield a discrepancy score.
Results and Discussion

Figure 26 shows the discrepancy between observed $d'$ values for sample A versus sample B as a function of the number of sample B trials, $n_B$; the parameter was true $d'$, and separate panels are shown for each value of $n_A$. There were two main findings. First, the discrepancy is slower to converge to 0 for larger values of true $d'$. Second, the number of sample B trials, $n_B$, required for the discrepancy to converge to 0 decreases with increasing absolute numbers of trial presentations; that is, larger numbers of observations are required to minimize the discrepancy between estimated values of $d'$ when larger ratios are investigated. Table 5 shows the maximum $n_A: n_B$ ratio that can be tolerated given each value of $n_A$ and each value of true $d'$. For example, for true $d' = 2.0$ and $A = 6$, the maximum $n_A: n_B$ ratio that can be tolerated is 1:1. However, for the same value of true $d'$ and $A = 480$, an $n_A: n_B$ ratio of 4.8:1 can be tolerated without risking an artificial difference in observed $d'$ between sample A and sample B. The subsequent experiments were modeled after Crum and Hafter (2008), and so 6:1 was the desirable ratio of expected (sample A) to unexpected (sample B) trials. Thus, going forward the target value of true $d'$ was 1.5, and the ratio of expected to unexpected trials was chosen to be 480:80.

<table>
<thead>
<tr>
<th>true $d'$</th>
<th>Number of sample A trials, $n_A$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>6</td>
</tr>
<tr>
<td>0.5</td>
<td>1.2:1</td>
</tr>
<tr>
<td>1.0</td>
<td>1:1</td>
</tr>
<tr>
<td>1.5</td>
<td>1:1</td>
</tr>
<tr>
<td>2.0</td>
<td>1:1</td>
</tr>
<tr>
<td>2.5</td>
<td>1:1</td>
</tr>
<tr>
<td>3.0</td>
<td>1:1</td>
</tr>
<tr>
<td>3.5</td>
<td>1:1</td>
</tr>
<tr>
<td>4.0</td>
<td>1:1</td>
</tr>
</tbody>
</table>

Table 5. Simulation 2. Maximum tolerable ratio of sample A to sample B trials, $n_A:n_B$ for each tested combination of $n_A$ and true $d'$. 
Summary

The lessons from Simulations 1 and 2 are as follows. First, differences between $d'$ estimates for sample A and sample B were larger for larger values of true $d'$. Second, this discrepancy was larger for smaller numbers of absolute trial presentations. Third, the ratio of sample A to sample B trials, $n_A: n_B$, can be increased with increasing numbers of trial presentations. Finally, the two corrections examined here, i.e., those by MacMillan and Creelman (2005) and Hautus (1995) were similar in terms of discrepancy between $d'$ estimates for sample A and sample B. On these bases, the following decisions were made about experimental design for Experiments 8-10. First, a relatively low value of true sensitivity was targeted, i.e., $d' = 1.5$. Prior to participating in any of the three experiments, listeners completed a single calibration session in which the amplitude of the to-be-detected pure-tone signal that corresponded to $d' = 1.5$ was estimated, and all stimuli were generated at that level (see Appendix D for calibration procedure). Second, a large number of trials was presented. Specifically, following Crum and Hafter (2008), the ratio of expected to unexpected trials was 6:1; thus, based on the results of Simulation 2, the absolute number of trials for expected and unexpected conditions was set to 480:80. Finally, where applicable, proportions of 0 and 1 were adjusted using the correction prescribed by MacMillan and Creelman (2005).
Figure 26. Simulation 2. Discrepancies between observed $d'$ for sample A and sample B ($d'_A - d'_B$) as a function of the number of sample B trials, $n_B$. Separate lines are shown for each value of true $d'$ and separate plots are for each value of the number of sample A trials, $n_A$. 
CHAPTER 6: FREQUENCY-TIME INTERDEPENDENCE IN ATTENTION

The experiments reported in this chapter examined attentional tuning to frequency-time locations based on trajectory information in continuous (Experiment 9) and discrete (Experiment 10) cues. The experiments were modeled after a study by Crum and Hafter (2008), where 60% of signals were presented at the expected frequency and 10% of signals belonged to each of four unexpected conditions. However, the results of Experiment 7 and simulations reported in Chapter 5 indicated that differences in observed sensitivity may obtain for conditions with unbalanced numbers of observations that do not reflect differences in true sensitivity. Thus, Experiment 8 was conducted to confirm the choice of experimental parameters selected to minimize the likelihood that heightened sensitivity to detect an expected (sample A) relative to an unexpected (sample B) frequency was due to an unbalanced number of observations in the to-be-compared samples.

Experiment 8

Method

Design

The design was a 2 (Analysis Type: unbalanced, balanced) x 2 (Trial Type: signal, noise) within-subjects factorial. Listeners detected a pure-tone signal in noise by responding ‘yes’ (signal present) or ‘no’ (signal absent); the signal was present on 50% of trials. As in Experiment 7, two analyses were conducted on each listener’s data set; first an unbalanced analysis assigned 60% and 10% of signal trials to two to-be-compared samples (sample A and sample B, respectively). A second, balanced analysis assigned 20% of signal trials to sample A and sample B. For both analyses, $d’$ was estimated for each sample as described below.

Participants
Eight (n = 3, female) listeners participated in return for monetary compensation equal to $10 / hour. All participants were determined to have normal hearing according to a standard audiometric screening (thresholds \( \leq 25 \) dB for frequencies of 250 Hz, 500 Hz, 1000 Hz, 2000 Hz, 4000 Hz, 8000 Hz)\(^8\). Listeners reported a range of formal musical training (1.5 – 10 years, \( M = 6.4 \) years, \( SD = 2.8 \)).

**Stimuli and apparatus**

Stimuli consisted of a 1000-ms segment of 61-dB noise that was followed by louder 65-dB noise burst of variable duration (175 ms, 250 ms, 325 ms). A 100-ms segment of quieter 61-dB followed the burst. In contrast to Experiment 7, where white noise was presented, in the current experiment noise was bandpass filtered between 200 Hz and 10,000 Hz. On 50% of trials a pure-tone signal was embedded in the final segment of noise. The frequency of the signal was drawn on each trial from a uniform distribution ranging between 850 Hz and 2150 Hz. For all signal trials, signal intensity was set equal to the intensity determined to yield \( d' = 1.5 \) for each individual subject; see Appendix D for details.

Stimuli were generated using MATLAB software (The Mathworks, Inc.). Stimulus generation and response collection were controlled using E-Prime 2.0.8.74 software (Psychology Software Tools, Inc.) running on a Dell Optiplex computer. Auditory stimuli were produced during testing by an internal SoundBlaster Audigy SE soundcard and presented over Sennheiser HD 414 headphones to listeners in a double-walled sound chamber. Responses were made using a serial response box.

**Procedure**

Listeners did not complete a familiarization block prior to the experiment; all listeners completed a calibration experiment (see Appendix D) in the session prior to beginning the
current experiment, and the task was identical. On each trial, listeners detected pure-tone signals following the variable-duration noise burst by responding ‘yes’ (signal present) or ‘no’ (signal absent). Signals were present on 50% of trials, and were always presented at the intensity determined for an individual to correspond to $d' = 1.5$.

Listeners completed 32 experimental blocks over several sessions completed on separate days; individual listeners completed between 3 and 19 blocks per session, and the duration of testing ranged between 30 minutes and 3 hours in a single session. Listeners completed 150 trials per block (75 signal trials, 75 noise trials). Overall, listeners completed a total of 4800 trials (2400 signal trials and 2400 noise trials), and on average, the experiment took approximately 5 – 6 hours.

**Data analysis**

The signal detection metric $d'$ quantified sensitivity to detect the pure-tone signals, and was calculated as in Experiment 7. An unbalanced and a balanced analysis were conducted separately for each listener’s data. For both analyses, only one third of the total trials were considered. For the unbalanced analysis, 60% and 10% of each listener’s signal trials were randomly assigned to two to-be-compared samples, referred to as sample A and sample B, respectively. For the current experiment, this meant that the number of sample A trials, $n_A$, was 480, and the number of sample B trials, $n_B$, was 80. For the balanced analysis, 20% and 20% of each listener’s signal trials were randomly assigned to two samples. For the balanced analysis, $n_A = n_B = 160$ signal trials. No trials were included in both sample A and sample B. Separate hit rates were calculated for each sample, but only a single false alarm rate was calculated across all noise trials (Werner, et al., 2009); the number of noise trials used in both analyses was 800.

**Results and Discussion**
Figure 27 shows observed $d'$ for sample A and sample B for the unbalanced and balanced analyses. It is clear from the figure that observed $d'$ values did not differ between samples according to either the balanced ($d'_{A} = 1.62 \pm .022; d'_{B} = 1.59 \pm 0.21$) or unbalanced analyses ($d'_{A} = 1.60 \pm .022; d'_{B} = 1.58 \pm 0.20$). This was confirmed by separate paired-samples $t$-tests (balanced: $t(7) = 0.21, p = 0.84$; unbalanced: $t(7) = -1.17, p = 0.28$). Thus, the design of the current experiment informed by the results of Simulations 1 and 2 was successful in eliminating the artifactual difference in estimated $d'$ between samples observed for the unbalanced analysis in Experiment 7.

![Figure 27. Experiment 8. Observed $d'$ values for sample A (white bars) and sample B (shaded bars) are shown for the balanced and unbalanced analyses. No differences in sensitivity were observed between samples for the balanced or unbalanced analysis.](image-url)
In the experiments that follow, the aim was to investigate differences in perceptual sensitivity between ‘expected’ and ‘unexpected’ conditions, extending the work of Crum and Hafter (2008). The paradigm was the same as in Experiment 8, with the exception that ascending or descending cues were presented during the 1000-ms segment of broadband noise. Cues were continuous tone glides in Experiment 9 and discrete tone sequences in Experiment 10.

Experiment 9

Overview

Experiment 9 extended Crum and Hafter’s (2008) study, in which listeners extrapolated the frequency-time trajectories of ascending tone glides to the frequency-time location at which a pure-tone signal could have been expected to occur; here, cues were ascending and descending tone glides. In the current study, listeners detected pure-tone signals in noise; signals were present on 50% of trials, and were either expected (60% of signal trials) or unexpected (40% of signal trials). Expected signals were defined based on extrapolation of the tone glide through a variable-duration noise burst. A schematic of stimuli is shown in Figure 28. Cues ascended (Panel A) or descended (Panel B) with constant velocity equal to 800 Hz/s. On each trial, the starting frequency of the cue was selected randomly from a uniform distribution, thus listeners could not use a strategy that involved tuning attention to a particular frequency region. Moreover, the noise burst separating the cue and the signal took on one of three durations from trial to trial, thus listeners could not learn a frequency relationship between the end of the glide and the expected signal. Instead, listeners had to use the velocity of the cue to project attention to the frequency-time location where the signal would occur. It was expected that if listeners were capable of making use of the cue’s velocity information, they would be most sensitive to detect signals at the expected frequency location and less sensitive to detect unexpected signals that
**expected signal frequency**

**Signal Trial (50%)**
Response: ‘yes’ (signal present)

**Noise Trial (50%)**
Response: ‘no’ (signal absent)

Noise Burst = 175ms, 250ms, 325ms
Stimulus and task diagram for Experiment 9. Listeners detected a pure-tone signal following a variable-duration noise burst by responding ‘yes’ (signal present) or ‘no’ (signal absent). Signals were present on 50% of trials (signal trials). Signals were preceded by an (A) ascending or (B) descending continuous tone glide cue that had constant velocity equal to 800 Hz/s. On 60% of signal trials, the signal occurred at the frequency location that was defined as ‘expected’ based on extrapolation of the cue through the noise burst. On the remaining 40% of signal trials, signals occurred at each of four unexpected frequency locations (±20%, 40% of the expected frequency). The signal, noise burst, and 1000 ms preceding the noise burst was presented against a background of low-level broadband noise to mask the signal.
were higher or lower than the expected signal. With respect to cue direction, based on an auditory gravity hypothesis (Henry & McAuley, 2009; Henry, et al., 2009), it was expected that listeners would overshoot the frequency location of expected signals for descending, but not ascending cues; that is, listeners were expected to listen too low for descending cues.

**Method**

**Design**

The design was a 2 (Trial Type: signal, noise) x 2 (Cue Direction: ascending, descending) x 3 (Noise Burst Duration: 175 ms, 250 ms, 325 ms) x 5 (Signal Frequency: 0%, ±20%, ±40%) mixed factorial. Cue Direction was a between-subjects factor, while Trial Type, Noise Burst Duration, and Signal Frequency were varied within subjects. Listeners detected pure-tone signals in noise by responding ‘yes’ (signal present) or ‘no’ (signal absent).

**Participants**

Six listeners were undergraduate and graduate students at a large Midwestern university, and were paid $10 / hour for their time. Four listeners also participated in Experiment 8. All participants were confirmed to have normal hearing according to a standard audiometric screening, and all were naïve to the purpose of the experiment. Listeners reported a range of formal musical training (1.5 – 14 years, $M = 7.3$ years, $SD = 4.2$ years). Three listeners heard ascending cues, and three listeners heard descending cues.

**Stimuli and apparatus**

A schematic of stimuli is shown in Figure 28. Stimuli consisted of a 60-dB, 1000-ms tone glide, ascending (Panel A) or descending (Panel B) in linear frequency space at a velocity of 800 Hz/s. Starting frequency of the glide was sampled on each trial from a uniform distribution; for ascending cues the distribution ranged between 476 Hz and 756 Hz, and for descending cues the
distribution ranged between 2476 Hz and 2756 Hz. The glide was followed by a 65-dB broadband noise burst (bandpass filtered between 200 Hz and 10,000 Hz) that lasted one of three durations: 175 ms, 250 ms, 325 ms. Following the noise burst, a 100-ms pure-tone signal was presented on 50% of trials (signal trials) at the amplitude determined to correspond to $d' = 1.5$ for an individual listener (see Appendix D). On 60% of signal trials, the signal occurred at the expected frequency based on extrapolation of the cue through the variable-duration noise burst. On 40% of signal trials, the signal was presented at one of four unexpected signal frequencies ($\pm 20\%$, $\pm 40\%$); that is, each unexpected signal frequency was presented on 10% of signal trials. The range of signal frequencies was between 850 Hz and 2150 Hz, as in Experiment 8. The whole stimulus (i.e., cue + noise burst + (signal)) was presented against a background of 61-dB broadband noise (200 Hz – 10,000 Hz).

Stimuli were generated using MATLAB software (The Mathworks, Inc.). Stimulus generation and response collection were controlled using E-Prime 2.0.8.74 software (Psychology Software Tools, Inc.) running on a Dell Optiplex computer. Auditory stimuli were produced during testing by an internal SoundBlaster Audigy SE soundcard and presented and were presented over Sennheiser HD 414 headphones to listeners in a double-walled sound chamber. Responses were made using a serial response box.

Procedure

Listeners heard recorded instructions, then completed an 18-trial familiarization block without feedback prior to the experiment. All listeners obtained PC equal to at least 94.4% (17 out of 18) during familiarization. On each trial, listeners detected pure-tone signals following the variable-duration noise burst by responding ‘yes’ (signal present) or ‘no’ (signal absent). Signals were present on 50% of trials, and for familiarization were presented at 60 dB.
Listeners completed 32 experimental blocks over several sessions completed on separate
days; individual listeners completed between 1 and 21 blocks per session, and the duration of
testing ranged between 15 minutes and 3 hours per session. Listeners completed 150 trials per
block, responding 25 times to each Trial Type x Noise Burst Duration combination. Overall,
each listener completed a total of 4800 trials (2400 signal trials and 2400 noise trials).

Data analysis

Perceptual sensitivity ($d'$) was estimated for each listener for each combination of Noise
Burst Duration and Signal Frequency; hits were defined as ‘yes’ responses to signal trials, and
false alarms were defined as ‘yes’ responses to noise trials. Separate hit rates were calculated for
each value of Signal Frequency, but only one false alarm rate was calculated across all noise
trials per noise burst duration.

With respect to the listening band, of primary interest here was the center frequency. To
quantify center frequency, a weighted measure (WM; Hubbard, 1995a) was calculated, which is
given by the formula:

$$ WM = \frac{(d'r_{-4}+d'r_{-2}+d'r_{0}+d'r_{+2}+d'r_{+4})}{d'r_{-4}+d'r_{-2}+d'r_{0}+d'r_{+2}+d'r_{+4}} $$

The WM score quantifies the degree and direction of the shift in listening band center relative to
the expected signal frequency. Negative WM values indicate an underestimation of expected
signal frequency (i.e., listening band center frequency is lower than expected signal frequency)
and positive values indicate an overestimation of expected signal frequency (i.e., listening band
center frequency is higher than expected signal frequency). WM values are reported as
proportions; for example, a WM value equal to -.15 would correspond to a 15% underestimation
of expected frequency.

Since data for only a small number of listeners were collected for each condition, Monte
Carlo simulations were conducted to replace standard significance tests. The procedure was as follows. First, the data for all listeners in a single between-subjects Cue Direction condition were pooled. Then, half of the data, ignoring listener, were selected for analysis. Note that this procedure is similar to Crum and Hafter (2008), with the exception that data for all listeners were pooled prior to splitting the sample in half; Crum and Hafter selected and analyzed half of each listener’s data. However, this modification was deemed necessary given the results reported in Chapter 5, which suggested that inspecting only half of one subject’s data could potentially inflate the benefit observed for detecting the expected signal relative to the unexpected signals.

Next, a $d'$ value was estimated for each combination of Noise Burst Duration and Signal Frequency. Estimated values of $d'$ were then used to estimate WM, as described above. Thus, a WM value was calculated for each Noise Burst Duration using half of the data collapsed across listeners; this process was repeated for a total of 100 iterations separately for the ascending and descending cues. Comparisons between conditions were made in a pair-wise fashion, and the proportion of times WM values for one condition exceeded WM values for a second condition was treated similarly to a $p$-value. For example, if WM values for sample A exceeded WM values for sample B on more than 95 out of 100 iterations, the difference was taken to be significant at $\alpha = 0.05$.

Results

Figure 29 shows $d'$ as a function of Signal Frequency for each Noise Burst Duration for individual listeners; data for ascending glides are shown on the left and for descending glides are shown on the right. In general, the results for ascending cues were consistent with the findings reported by Crum and Hafter (2008); values of $d'$ were highest for the expected signal frequency, and fell off with increasing frequency distance from the expected location. One exception is the
Figure 29. Experiment 9. Perceptual sensitivity ($d'$) is shown as a function of Signal Frequency separately for each Noise Burst Duration. Results for ascending cues are shown on the left for individual listeners. Sensitivity was highest at the expected signal frequency, and fell with increasing frequency distance from the expected location. Results for descending cues are shown on the right for individual listeners. Sensitivity was highest at or below the expected frequency, and was close to chance for signal frequencies above the expected location.
-40% signal frequency, which was consistently better detected than the -20% for listeners CC and KB. In contrast, for descending cues, sensitivity was highest for relatively low values of signal frequency; that is, signals presented at or lower than the expected frequency were better detected than relatively high signal frequencies, indicating that listeners were tuning in too low relative to the expected signal frequency. Listener DT also showed a tendency for heightened sensitivity to detect the expected signal over and above the shift of the listening band to the lower frequency range. One striking aspect of the data is that, despite signal amplitude being calibrated identically for the ascending and descending cue conditions, overall sensitivity was much higher for descending than for ascending cues; mean $d'$ values were 0.96 ($\pm 0.34$) and 1.95 ($\pm 0.28$), respectively. Finally, there were no systematic differences between the shapes of the listening bands as a function of Noise Burst Duration.

WM values are shown in Figure 30 for ascending and descending tone glide cues; the figure shows both values averaged across listeners and averaged across iterations of the Monte Carlo simulations. As expected, WM values were close to zero for ascending cues and negative for descending cues. Monte Carlo simulations indicated that WM values for ascending cues did not differ significantly from 0, regardless of Noise Burst Duration. That is, the listening band was centered on the expected signal frequency based on extrapolation of an ascending tone glide cue through the noise burst ($WM = 0.009$); despite the variable duration of the noise burst, listeners were capable of using velocity information to tune attention to the expected frequency location at the right time. There were no pairwise differences between WM values for the 175-ms, 250-ms, and 325-ms noise bursts. For descending cues, WM values were significantly below 0 for all three noise burst durations. Overall, listening bands were centered on a frequency 12% lower than the expected frequency based on extrapolation through the noise burst. Pairwise
Figure 31. Weighted measure (WM) scores for each Noise Burst Duration for ascending and descending cues. WM scores are shown both averaged over listeners (± SEM) and averaged over iterations of the Monte Carlo simulation conducted to test for significance (X’s).

comparisons revealed a significant difference between WM values for the 175-ms noise burst ($WM = -0.13$) and 250-ms noise burst ($WM = -0.11$).

Discussion

In the current experiment, listeners detected expected and unexpected pure-tone signals in noise, where expected signals were defined based on extrapolation of the frequency-time trajectory of an ascending or descending tone glide cue preceding the signal. The results for ascending tone glides replicated the main findings reported by Crum and Hafter (2008). Specifically, listeners were most sensitive to detect signals that occurred at the expected
frequency-time location based on extrapolation of the glide through a variable-duration noise burst. Critically, the starting frequency of the tone glide was randomly selected from a uniform distribution on each trial, and extrapolation duration (i.e., Noise Burst Duration) also varied from trial to trial, taking on one of three values (175 ms, 250 ms, 325 ms). Thus listeners could not adopt a strategy that involved tuning into a narrow range of frequencies or learning a frequency relationship between the end of the glide and the signal. Instead, the results indicate that listeners used the frequency-time trajectory of the tone glide cue to project attention to the future frequency-time location at which the signal was expected and thus better perceive a to-be-detected signal.

In contrast to the results for ascending cues, for descending glides, listeners were most sensitive to detect signals that were at or below the frequency that would have been expected based on extrapolation of the cue through the noise burst. That is, the center frequency of the listening band was lower than the expected signal frequency by 12%. This is contrary to the hypothesis put forward by Crum and Hafter (2008), who predicted that for descending cues, listeners should listen at or higher than the expected frequency, particularly when the noise burst was long. Instead, the current results are more consistent with an auditory gravity hypothesis (Henry & McAuley, 2009; Henry, et al., 2009), which suggests that listeners have internalized latent knowledge of gravity, and make use of a gravity-like principle when listening. An alternative suggestion regarding this finding comes from work indicating that producing descending pitch trajectories is easier and can be accomplished more quickly than producing ascending pitch trajectories (Sundberg, 1973; Xu & Sun, 2000, 2001). This explanation suggests an embodied aspect of auditory attention, whereby attention to auditory events is informed by how those sounds are produced by the human voice.
In the current experiment, there was a large difference in detectability for signals following ascending versus descending cues, despite calibration to equivalent difficulty levels. Specifically, $d'$ values were much higher overall for descending relative to ascending cues, especially at the signal frequencies at or below that expected based on extrapolation of the cue through the noise burst. One possible explanation for this finding is that this difference too is attributable to the ease with which pitch excursions produced by the voice can be executed in an ascending or a descending direction; perhaps attentional tracking and extrapolation is overall easier in a descending direction, leading to a boost in observed sensitivity for signals following a descending cue. These points will be more fully elaborated in the General Discussion.

Experiment 10 extended the results of the current study to discrete tone sequences cues. Listeners detected pure-tone signals in noise that were either expected or unexpected based on extrapolation of the frequency-time trajectory of a discrete tone sequence cue through the noise burst.

Experiment 10

Overview

Experiment 10 extended Experiment 9 to discrete tone sequences cues. Listeners detected pure-tone signals in noise; signals were present on 50% of trials, and were either expected (60% of signal trials) or unexpected (40% of signal trials). Expected signals were defined based on extrapolation of an ascending or descending tone sequence cue through a variable-duration noise burst. A schematic of stimuli is shown in Figure 31. Cues ascended (Panel A) or descended (Panel B) with constant velocity equal to 800 Hz/s. As in Experiment 9, the starting frequency of the cue was selected randomly from a uniform distribution on each trial, and the noise burst separating the cue and the signal took on one of three durations from trial to trial (i.e., 175 ms,
**expected signal frequency**

**Signal Trial (50%)**
Response: ‘yes’ (signal present)

Noise Burst = 175ms, 250ms, 325ms

-20%
-40%
0%
+20%
+40%

**Noise Trial (50%)**
Response: ‘no’ (signal absent)

Noise Burst = 175ms, 250ms, 325ms

-20%
-40%
0%
+20%
+40%
Figure 31. Stimulus and task diagram for Experiment 10. Listeners detected a pure-tone signal following a variable-duration noise burst by responding ‘yes’ (signal present) or ‘no’ (signal absent). Signals were present on 50% of trials (signal trials). Signals were preceded by an (A) ascending or (B) descending discrete tone sequence cue that had constant velocity equal to 800 Hz/s. On 60% of signal trials, the signal occurred at the frequency location that was defined as ‘expected’ based on extrapolation of the cue through the noise burst. On the remaining 40% of signal trials, signals occurred at each of four unexpected frequency locations (± 20%, 40% of the expected frequency). The signal, noise burst, and 1000 ms preceding the noise burst was presented against a background of low-level broadband noise to mask the signal.
205 ms, 325 ms). The SOA between individual sequence elements was fixed at 250 ms. Thus signals occurring following the 175-ms, 250-ms, and 325-ms noise burst durations were early, on time, and late, respectively with respect to the rhythm marked out by the cue.

Two hypotheses were contrasted. First, if the discrete nature of the stimuli promoted selective attention to the dimensions of frequency and time, it was expected that only the timing of the signals would contribute to detectability. Specifically, on-time signals were predicted to be better detected than early or late signals, regardless of signal frequency. No specific predictions were made regarding cue direction (i.e., ascending versus descending) on this view. On the other hand, if listeners attended jointly to frequency and time information in tone sequence cues, it was expected that as in Experiment 9, they would be most sensitive to detect signals at the expected frequency location, which varied with noise burst duration. Moreover, with respect to cue direction, it was expected that listeners would overshoot the frequency location of expected signals for descending, but not ascending cues, in line with an auditory gravity hypothesis (Henry & McAuley, 2009; Henry, et al., 2009); that is, the center frequency of the listening band was expected to be relatively low for descending sequence cues but close to the expected signal frequency for ascending sequence cues.

**Method**

**Design**

The design was a 2 (Trial Type: signal, noise) x 2 (Cue Direction: ascending, descending) x 3 (Noise Burst Duration: 175 ms, 250 ms, 325 ms) x 5 (Signal Frequency: 0%, ±20%, ±40%) mixed factorial. Cue Direction was a between-subjects factor, while Trial Type, Noise Burst Duration, and Signal Frequency were varied within subjects. Listeners detected pure-tone signals in noise by responding ‘yes’ (signal present) or ‘no’ (signal absent).
Participants

Six listeners were undergraduate and graduate students at a large Midwestern university, and were paid $10 / hour for their time. Four listeners participated in Experiment 8, and four listeners participated in Experiment 9. All participants were confirmed to have normal hearing according to a standard audiometric screening, and all were naïve to the purpose of the experiment. Listeners reported a range of formal musical training (2 – 9 years, $M = 6.3$ years, $SD = 2.5$ years). Three listeners heard ascending cues, and three listeners heard descending cues.

Stimuli and apparatus

A schematic of stimuli is shown in Figure 31. Stimuli consisted of a 60-dB five-tone sequence, ascending (Panel A) or descending (Panel B) in linear frequency space at a velocity of 800 Hz/s. The total duration of the sequence from the onset of the first tone to the onset of the final tone was 1000 ms; thus the SOA between individual tone onsets was 250 ms. The total frequency change of the sequence was 800 ms. Starting frequency of the sequence was sampled on each trial from a uniform distribution; for ascending cues the distribution ranged between 476 Hz and 756 Hz, and for descending cues the distribution ranged between 2476 Hz and 2756 Hz. A 65-dB, variable-duration (175 ms, 250 ms, 325 ms) broadband noise burst (200 Hz – 10,000 Hz) started concurrently with the onset of the final sequence tone. This meant that signals occurring after the 175-ms, 250-ms, and 325-ms noise bursts were early, on time, and late, respectively, with respect to the rhythm marked by the cue. Following the noise burst, a 100-ms pure-tone signal was presented on 50% of trials (signal trials) at the amplitude determined to correspond to $d' = 1.5$ (see Appendix D). As in Experiment 9, on 60% of signal trials, the signal occurred at the expected frequency based on extrapolation of the cue through the variable-duration noise burst. On 40% of signal trials, the signal was presented at one of four unexpected
frequencies (±20%, ±40%); that is, each unexpected signal frequency was presented on 10% of signal trials. The range of signal frequencies was between 850 Hz and 2150 Hz. The whole stimulus (i.e., cue + noise burst + (signal)) was presented against a background of 61-dB broadband noise (200 Hz – 10,000 Hz).

Stimuli generation, presentation, and response collection were identical to Experiment 9.

Procedure and data analysis

The procedure and data analysis were identical to Experiment 9. Listeners completed between 3 and 14 blocks per session, with sessions lasting between 30 minutes and 2.5 hours.

Results

Figure 32 shows $d'$ as a function of Signal Frequency for each Noise Burst Duration for individual listeners; data for ascending sequences are shown on the left and for descending sequences are shown on the right. In general, the results for ascending sequence cues were consistent with the results of Experiment 9 and the findings reported by Crum and Hafter (2008). In particular, listener PY showed the prototypical result, with increased sensitivity to detect the expected signal frequency relative to the unexpected signal frequencies, with the exception of the -40% signal frequency. However, for discrete tone sequences, there was a tendency for sensitivity to be highest for signal frequencies somewhat higher than the expected value; this was true for listeners KS and DT. For descending sequence cues, all listeners showed a pattern of results similar to that obtained for descending tone glides in Experiment 9. Specifically, listeners were most sensitive to detect signals occurring at or below the frequency expected based on extrapolation of the tone sequence through the noise burst. Listener HS showed overall best sensitivity to signals occurring at -20% of the expected value; listener KB
Figure 32. Experiment 10. Perceptual sensitivity ($d'$) is shown as a function of Signal Frequency separately for each Noise Burst Duration. Results for ascending cues are shown on the left for individual listeners. Sensitivity was highest slightly above the expected signal frequency. Results for descending cues are shown on the right for individual listeners. Sensitivity was highest at or below the expected frequency, and was close to chance for signal frequencies above the expected location.
showed a pattern of results similar to that exhibited by DT in Experiment 9, where in addition to a tendency to overshoot the expected signal location following a descending sequence, a local peak in sensitivity was observed for the expected signal frequency. In the current experiment, in contrast to the results of Experiment 9, there was not such a striking difference between overall sensitivity levels for ascending ($d' = 1.62 \pm 0.13$) relative to descending cues ($d' = 1.72 \pm 0.23$).

WM values are shown in Figure 33 for ascending and descending tone sequence cues; the figure shows both values averaged across listeners and averaged across iterations of the Monte Carlo simulations. Monte Carlo simulations revealed that WM values for ascending cues were significantly greater than zero and WM values for descending cues were less than zero, regardless of Noise Burst Duration. Listening bands were centered, on average, 3.6% higher than the expected frequency for ascending cues and 12.1% lower than the expected frequency for descending cues. Pairwise comparisons between Noise Burst Durations revealed that, for both ascending and descending cues, WM magnitude was smaller for the 325-ms noise burst ($WM_{asc} = 0.02; WM_{desc} = -0.10$) than for the 175-ms noise burst ($WM_{asc} = 0.04; WM_{desc} = -0.13$) or the 250-ms noise burst ($WM_{asc} = 0.04; WM_{desc} = -0.13$).

**Discussion**

In the current experiment, listeners detected expected and unexpected pure-tone signals in noise following discrete tone sequence cues. Expected signals were defined based on extrapolation of the trajectory implied by the tone sequence through a variable-duration noise burst. For ascending cues, listeners were most sensitive to detect signals that occurred somewhat higher than the expected signal frequency; the center frequency of the listening band was estimated to be 3.6% higher than the expected signal frequency. For descending sequences, listeners were most sensitive to detect signals that were at or below the frequency that would
have been expected based on extrapolation of the cue through the noise burst; the center frequency of the listening band was estimated to occur 12% lower than the expected frequency.

The systematic overshoot of the expected signal frequency for descending sequences cues was consistent with an auditory gravity hypothesis, which assumes that listeners apply internalized knowledge of gravity to auditory stimuli, attributing acceleration to descending auditory trajectories. However, the overestimation of expected signal frequency for ascending sequences cues was an unexpected result which was not predicted by either Crum and Hafter’s
temporal integration window hypothesis or an auditory gravity hypothesis. This point is returned
to in the General Discussion.

Pairwise comparisons for WM values across noise burst durations revealed that for both
ascending and descending tone sequence cues, the magnitude of the shift in listening band center
frequency was smaller for the 325-ms noise burst than for the 175-ms or 250-ms noise bursts.
This finding is consistent with the proposal put forth by Crum and Hafter (2008) regarding the
integration of velocity information across a sliding window. Based on this hypothesis, at the
termination of the cue, the estimate of velocity goes to zero; including a zero-velocity segment in
the integration window necessarily reduces estimated velocity. Thus the expected frequency is
farther along the extrapolated trajectory than the center frequency of the listening band based on
estimated velocity. Moreover, the longer the duration of cue off-time, the more severe the
underestimation of cue velocity, and the more pronounced the mismatch between expected signal
frequency and listening band center frequency.

General Discussion

The current experiments investigated detection for pure-tone signals that were either
expected or unexpected based on extrapolation of continuous or discrete cues through a variable-
duration noise burst. In Experiment 9, cues were continuous tone glides that either ascended or
descended at a constant velocity of 800 Hz/s. In Experiment 10, cues were discrete tone
sequences that either ascended or descended and implied the same velocity as in Experiment 9.
In both experiments, cue starting frequency was randomized from trial to trial, and extrapolation
duration (i.e., noise burst duration) took on one of three values: 175 ms, 250 ms, 325 ms. Thus
tuning attention to the frequency-time location of the expected signal required extrapolation of
the cue through the noise burst, and could not be accomplished by adopting a strategy that
involved either monitoring a narrow frequency region or learning a frequency relationship between the final frequency of the cue and the expected signal.

For ascending glide cues, the center frequency of the listening band, as indexed by WM, was not significantly different from the expected frequency location based on extrapolation of the cue through the noise burst. That is, sensitivity to detect the signal at the expected frequency was better than to detect unexpected frequencies higher or lower than expected. This result replicated the main findings of Crum and Hafter (2008). For ascending tone sequence cues, WM values indicated that listening bands were centered approximately 4% higher than the expected frequency based on extrapolation through the noise burst. Monte Carlo simulations revealed that this value differed significantly from the WM value for ascending glide cues. For descending glide and sequence cues, WM values indicated that the center frequency of the listening band was 12% lower than the expected frequency based on extrapolation through the noise burst. Overall, the large overshoot of the expected signal frequency location was consistent with an auditory gravity hypothesis (Henry & McAuley, 2009; Henry, et al., 2009).

Continuous versus discrete cues

The results for continuous tone glide cues (Experiment 9) and discrete tone sequences cues (Experiment 10) were quite similar and were consistent with listeners extrapolating the frequency-time trajectory of both cue types to tune attention to the expected signal frequency. This is consistent with the auditory motion hypothesis, which proposes that listeners are sensitive to velocity information in sounds that move continuously in frequency/pitch space and in discrete sounds that only imply motion. This finding is inconsistent with listeners selectively attending to the time and frequency dimensions. Indeed, no strong effects of Noise Burst Duration were observed that would suggest disruption of attentional tuning by signals occurring
either early or late with respect to the cue’s rhythm.

One difference that obtained between continuous and discrete cues was a slight overshoot of the expected frequency for ascending sequence cues relative to ascending glide cues. A similar effect has been observed in vision by Kerzel (2003), who compared representational momentum effects in vision for stimuli undergoing smooth motion versus stimulus sequences that implied motion. In general, Kerzel found much larger forward displacements for implied motion stimuli relative to smooth motion stimuli. Moreover, he parametrically varied the SOA between sequence elements such that motion became less good with increasing SOA, and observed increasing magnitudes of forward displacement with increasing SOAs. He suggested that a likely reason for this difference was that automatic extrapolation is carried out by observers when viewing a discrete stimulus implying motion. Even though the task of the observer is to judge the position of the final sequence element, attention is automatically deployed to the next spatial position given sequence velocity; adjustment back towards the to-be-judged element is incomplete, leading to a forward displacement in memory. Kerzel argued that extrapolation is not necessary for continuously moving stimuli, as all motion information is contained in the stimulus, so forward displacement is not typically observed to the same degree.

Providing a potential bridge for this hypothesis to the auditory stimuli, Hubbard (1995a) found a larger degree of forward displacement for the probe element following a discrete tone sequence than following a continuous tone glide.

*Ascending versus descending cues*

The large differences observed for ascending versus descending cues in the current studies deserve special attention. Specifically, the center frequency of the listening band overshot the expected signal frequency to a large extent for descending cues, but not for ascending cues.
The magnitude of the mismatch between the listening band and the expected signal frequency for descending cues was 12%, which corresponds to a difference of approximately 2 ST.

It is suggested that listeners made use of an auditory gravity heuristic which caused descending cues to be heard as accelerating (Henry & McAuley, 2009; Henry, et al., 2009); thus listeners centered their listening bands lower than the frequency defined as expected based on constant-velocity extrapolation. On this view, listeners have internalized knowledge of environmental invariants present in their environments, e.g., gravity. This is in line with the view of Shepard (1984), who suggested that all cognitive activities – including perceiving, imagining, and dreaming – are subject to the constraints of internalized environmental regularities. In vision, several studies support the influence of an internalized knowledge of gravity on perception and memory. For example, Freyd, Pantzer, and Cheng (1988) presented participants with a drawing of a flower pot either hanging from a hook or sitting on a table top. Then, memory for the spatial location of the pot was tested using drawings with the hook or table removed. In both cases, memory for the pot’s location was displaced downward, consistent with gravitational effects. Hubbard and Bharucha (1988) tested displacement in memory for the position of a visual stimulus that underwent apparent motion in either a horizontal or a vertical direction. The magnitude of displacement of vertically-moving stimuli was larger for downward relative to upward trajectories; moreover, horizontally-moving stimuli were also displaced downward in memory, again consistent with an internalized gravity hypothesis.

The current results are consistent with several previous studies that are suggestive of an auditory analogue of an internalized gravitational force. First, Henry and McAuley (2009) and Henry et al. (2009) have shown asymmetries in the magnitude of the auditory kappa and tau effects, such that larger effects were observed for descending trajectories that were potentially on
the border of the auditory motion range; the authors suggested that descending sequences on the
skirts of the auditory motion range were likely heard as accelerating, and so were pushed into a
range in which discrete sequences were heard in terms of implied motion, more so than for their
ascending counterparts. Second, Hubbard (1995a) tested representational momentum for
ascending and descending tone sequences and glides, and found a tendency for forward
displacement of the final tone in memory to be larger for descending relative to ascending
stimuli. Finally, Larson and VanHandel (2005) proposed that environmental forces such as
gravity, intertia, and magnetism operate on listener expectations in musical contexts. They
showed that listeners judged three-tone sequences as sounding more complete when the final
tone descended, in line with a gravitational hypothesis.

Although the results of the current studies and those reviewed here are consistent with an
auditory gravity hypothesis, it is unclear what adaptive significance would be held by an
internalized gravity heuristic operating in the auditory modality. One possible means for this link
is the mapping of auditory pitch onto a visual spatial representation (e.g., Douglas & Bilkey,
2007; Lidji, et al., 2007; Woods, Alain, Diaz, Rhodes, & Ogawa, 2001). In a study conducted by
Eitan and Granot (2004a, 2004b), listeners were asked to imagine the motion of an imagined
character while listening to sounds undergoing musical transformations such pitch excursions
and crescendos (i.e., a gradual increase in volume). They found that pitch direction was
associated with changes in the vertical spatial position of the imagined character in an
asymmetrical way, such that descending pitch changes were strongly associated with downward
motion, but ascending pitch changes were only very weakly associated with upward motion,
consistent with forces of gravity acting on the imagined character. The idea that musical
parameters such as pitch direction have consistent correlates with motion in the physical world
suggests a potential mechanism by which internalized physical forces such as gravity could have come to be applied to auditory stimuli.

A perhaps simpler explanation for the observed directional effects concerns the use of a linear frequency scale to construct the cues in the current study. There has been a long-standing debate in pitch perception literature regarding the appropriate psychological scale for the physical dimension of frequency (Burns, 1999). A number of studies provide evidence that a logarithmic frequency scale (i.e., a musical scale where equal steps correspond to ST) is the psychologically relevant scale for the manipulation of frequency (Attneave & Olson, 1971). Here, cues were used that traversed frequency space linearly. Thus, the descending cues presented in the current studies mapped onto a logarithmic frequency space would be heard as accelerating, while ascending cues would be heard as decelerating. This explanation would account for the overshoot of the expected signal frequency following descending cues, but would also predict an undershoot of expected signal frequency for ascending cues, which was not observed here.

Another possibility is suggested by the observation that, in classical, pop, and ‘ethno’ music, ascending pitch intervals tend to be larger than descending intervals (Vos & Troost, 1989). Experience with frequently-heard descending intervals that tend to be smaller than frequently-heard ascending intervals may lead a listener to overestimate the frequency distance of a descending excursion relative to its ascending counterpart (Russo & Thompson, 2005). In the current study, overestimating frequency change would mean an overestimation of cue velocity, and thus an overshoot of the expected signal frequency. This is a viable possibility that deserves additional attention.

A final alternative explanation for the difference between ascending and descending cues
Auditory Motion for Continuous and Discrete Sounds

has to do with the ease with which ascending versus descending pitch trajectories can be produced by the human voice\textsuperscript{10}. Several studies examining the maximum speed with which variable-size pitch intervals can be produced have demonstrated that descending pitch intervals can be produced more quickly than ascending pitch intervals of the same size (Sundberg, 1973; Xu & Sun, 2000, 2001); however, this difference was eliminated in a sample of trained singers (Sundberg, 1973), suggesting that the natural mismatch in the ease with which ascending versus descending pitch excursions can be performed can be overridden by practice. This hypothesis suggests an embodied interpretation of the current results, in that differences in attentional tuning by ascending versus descending cues are attributed to a latent knowledge of one’s own production capabilities. This hypothesis affords a number of tests based on individual differences. For example, trained singers should not show an asymmetry between ascending and descending cues to the same extent as individuals without voice training. Moreover, Xu and Sun (2001) observed that female participants required less time in the acceleration and deceleration phases of the pitch interval production, presumably because their smaller laryngeal mass builds up less laryngeal inertia. Thus it could be hypothesized that gender differences should be observable in the present paradigm in terms of the amount of overshoot of the expected signal frequency.

*Parallels to visual extrapolation and the role of the noise burst*

Overall, the results of the current set of experiments are consistent with the hypothesis that listeners make use of frequency-time trajectory information that is either made explicit by continuous glide cues or implied by discrete sequences cues. At this point, it is useful to make an analogy to several studies in vision which posed a similar task to observers. In a study by Churchland et al. (2003), eye movements of monkey observers were monitored while the
monkeys tracked a visual stimulus that moved along trajectories that differed in the type of motion implied by the stimulus. In ‘normal’ motion conditions, the stimulus underwent continuous, uninterrupted motion along a horizontal path. In ‘occluded’ motion conditions, the stimulus followed the same path as in the normal condition, but disappeared transiently behind an occluding object with variable size and position. Finally, in a ‘blinked’ condition, the stimulus transiently disappeared for the same distance and duration as in the ‘occluded’ conditions, but no occluding object was present. Overall, there was a decrease in eye velocity at the time of target disappearance for both the occluded and blinked conditions relative to the normal conditions. However, the decrease in eye velocity was much larger for the blinked than for the occluded conditions, and in fact, eye velocity profiles for the occluded conditions for several monkeys more closely resembled the normal profiles than the blinked profiles. That is, pursuit of a visual target was sustained when an occluder was present which justified target disappearance. A study by Yantis (1995) with human observers similarly demonstrated the importance of the presence of an occluding object for maintaining a constant object representation through a period where a visual target was not visible.

In audition, the presence of an occluding object (i.e., a noise burst) also seems to be important for transiently missing information to be perceived as present. For example, perceptual restoration refers to a phenomenon whereby a sound that is interrupted by a noise burst is heard as continuing through the noise. Restoration occurs in both tonal and speech stimuli (Warren, 2008). There are two conditions prerequisite for hearing the interrupted sound continue through the noise burst. First, the frequency range of the noise burst must overlap with the frequencies of the stimulus to be filled in. Second, the noise burst must be of sufficient amplitude to mask the interrupted sound. Indeed, Crum and Hafter (2008) reported that, in their study, the experience
was of the cue continuing through the noise burst. However, several studies have varied the presence and intensity of noise during a gap through which listeners extrapolated gliding sounds, and found that although masking noise is important for the experience of continuity, its absence does not impact extrapolation based on a trajectory (Dannenbring, 1976; Kluender & Jenison, 1992; Tougas & Bregman, 1990). Here, although the frequency range of the noise burst overlapped the frequencies of the cues and signals, the noise burst was not sufficiently loud to mask the cue, and there was not sufficient evidence that the cue continued through the noise burst. Thus it can be concluded that the current results did not depend on the experience of perceptual continuity. However, it is noted that the role of the noise burst in the current paradigm is readily testable, and manipulating the nature of the noise burst in a similar paradigm has the potential to shed additional light on the mechanisms that underlie the projection of auditory attention to future frequency-time locations.

Summary and Conclusions

Three experiments demonstrated that listeners use frequency-time trajectory information to project attention to a future frequency-time location at which a to-be-detected signal is expected; listeners used both explicit trajectory information conveyed by continuous tone glides and implied trajectory information conveyed by discrete tone sequences to tune attention. A large asymmetry was observed between ascending and descending cues that was consistent across cue type; listening band center frequency occurred very near to the expected signal frequency for ascending cues, but listeners overshot the expected signal frequency following descending cues by approximately 12%. This overshoot of the signal frequency following descending cues was proposed to be the result of an application of an auditory gravity-like expectation that resulted in descending cues being heard as accelerating; however, several alternative explanations were
suggested, including the possibility that attentional tracking of auditory stimuli is influenced by the ease with which the to-be-tracked pitch trajectory can be produced by the human voice. In what follows, Part III of this dissertation provides an integrated discussion of the findings from the experiments reported in Parts I and II and summarizes the main contributions of this research.
PART III
CHAPTER 7: CONCLUSIONS AND IMPLICATIONS

In this chapter, I will summarize the results from the experiments reported in this dissertation. I then will turn to an integrated discussion of main findings placed in the context of previous research. Finally, theoretical implications that follow from the work are discussed; potential avenues for future research are highlighted throughout.

Summary of Results

In Part I, six experiments examined judgments of duration and pitch-change of a variable-velocity comparison stimulus relative to a constant-velocity standard stimulus in order to test the predictions of an auditory motion hypothesis. In Experiments 1 – 3, standard and comparison stimuli were tone glides that moved continuously in pitch space, whereas in Experiments 4 – 6, stimuli were discrete tone sequences that implied motion in pitch space. When stimuli were tone glides, results were consistent with the auditory motion hypothesis. Specifically, comparison duration was overestimated when comparison velocity was greater than standard velocity; however, comparison duration was underestimated when the comparison glide was slower than the standard. On the other hand, comparison pitch-change was overestimated when the comparison was slower than the standard and underestimated when the comparison was faster than the standard. When stimuli were discrete tone sequences, no systematic distortions were observed in perceived duration or pitch-change of the comparison. In terms of the auditory motion hypothesis, this finding suggests that velocity information was less clearly conveyed, or easier to ignore, for discrete sequences that implied motion in pitch space.

In Part II, a methodological issue was explored that arises when perceptual sensitivity is compared between expected and unexpected conditions, where the expected condition is the more frequently presented condition. The main contribution of this examination was the finding
that biases in estimation of $d'$ can lead to an observed boost in sensitivity for the ‘expected’ condition relative to the ‘unexpected’ condition that does not reflect differences in true sensitivity, in line with the report of Miller (1996). However, the results of two simulations reported here revealed that targeting a low value of true $d'$ and collecting a large number of observations can work to prevent this measurement artifact, and thus strengthen conclusions regarding differences in detection performance for expected versus unexpected events. Finally, the results of Experiment 8 confirmed that this measurement artifact was eliminated when choice of experimental parameters was informed by the simulation results.

Two experiments examined signal detection performance for pure-tone signals that occurred at expected or unexpected frequency-time locations, where the expected location was defined based on extrapolation of an ascending or descending cue through a variable-duration noise burst in order to test the hypothesis that listeners use pitch-time trajectory information to tune attention to the anticipated location of a to-be-detected signal. In Experiment 9, cues were continuous tone glides, and in Experiment 10, cues were discrete tone sequences. Consistent with the auditory motion hypothesis, results were similar for continuous and discrete cues and indicated that listeners made use of trajectory information to tune attention in frequency and time. Specifically, for ascending cues of both types, listeners showed enhanced sensitivity to detect signals that occurred at the expected frequency-time location based on extrapolation of the cue trajectory. For continuous and discrete descending cues, listeners showed a shift in listening band center frequency to a lower frequency range than the expected signal location. That is, listeners overshot the expected signal frequency by 12%, or approximately 2 ST, consistent with an auditory gravity hypothesis.

There are four primary contributions of this work. First, continuous and discrete auditory
patterns were shown to differ in the degree to which they conveyed velocity information in Part I, with discrete sequences implying velocity information that was either less clear or easier to ignore than velocity information conveyed by continuous auditory stimuli moving in pitch space. Second, listeners were shown to similarly tune attention on the basis of trajectory information conveyed by continuous and discrete auditory patterns in Part II. Possible reasons for differences in results for continuous versus discrete sounds in Parts I and II are discussed below. Third, descending auditory stimuli were observed to result in an overshoot of the expected frequency location of a to-be-detected signal relative to ascending cues, in line with an auditory gravity hypothesis. Fourth, caution is warranted in comparisons of sensitivity to detect expected and unexpected signals when expectation is defined based on numbers of observations; targeting a low value of true sensitivity and collecting a large number of observations can combat measurement artifacts. In what follows, these contributions will be discussed in the context of previous work and current theory, and directions for future research will be suggested.

Part I: Continuous versus discrete stimuli

In Part I, results were consistent with the auditory motion hypothesis when stimuli moved continuously in pitch space (Experiments 1 – 3), but the strong effects observed for continuously moving stimuli were absent when stimuli were discrete sequences that only implied motion in pitch space (Experiments 4 – 6). Stimuli were matched on the dimensions of total duration, total pitch change, and velocity. One caveat is that there was a mismatch in velocities examined in Experiments 1 – 3 and those examined in Experiments 4 – 6. However, a follow-up experiment with matched velocities indicated that differences in velocities for the continuous and discrete stimuli does not account for behavioral differences observed between the two conditions. Behavioral differences between continuous and discrete conditions suggest that velocity
information was conveyed less clearly, or was alternatively easier to ignore, for discrete tone sequences than for continuous tone glides. This is consistent with several previous studies that have directly compared performance with matched continuous and discrete auditory stimuli.

Henry and McAuley (in press) asked listeners to provide velocity estimates for continuous tone glides and discrete tone sequences that ascended or descended in pitch; continuous and discrete stimuli were matched in terms of total duration, total pitch change, and velocity. Velocity was manipulated by either holding duration constant while varying pitch-change or holding pitch-change constant while varying duration. The authors observed that, for continuous tone glides, listeners made approximately equal use of pitch change and duration information when estimating velocity; however, for discrete tone sequences, listeners made use of duration exclusively, and velocity estimates showed no dependence on tone sequence pitch change. Similarly, Strybel et al. (1998) examined perceived velocity of sounds moving in auditory physical space, and found that for discrete sequences of sounds that created a sensation of auditory apparent motion, perceived velocity depended on the timing, but not the spacing, of individual sequence elements. Although Strybel et al. did not directly compare these results to perceived velocity of continuously moving stimuli, results from previous studies indicate that perceived velocity would likely have depended on both spatial and temporal properties of the stimuli (Perrott, Buck, Waugh, & Strybel, 1979; Waugh, Strybel, & Perrott, 1979).

An additional study that directly compared continuous and discrete auditory stimuli was conducted by Crum and Bregman (2006). In this study, listeners heard sounds that changed continuously in timbre (i.e., relative intensity of harmonics was modulated continuously), and indicated by button-press when they first noticed the change. Continuous and discrete stimuli were identical in terms of rate of timbre change; discrete stimuli were created by removing
segments of the continuous stimulus and replacing them with silence. Listeners were significantly faster to detect the timbre changes in discrete stimuli than continuously changing stimuli.

In all of these studies, it is important that the discrete stimuli were isochronous. Indeed, one possibility is that the authors would not have obtained the same results with irregularly timed discrete stimuli. The rationale for this proposal comes from Jones’ dynamic attending theory (Jones, 1976, 2004), in which attention is conceptualized as rhythmic and flexible (McAuley, 1995), able to be entrained by rhythmic environmental events; the result is that attentional oscillations peak at the anticipated onset times of events making up a discrete sequence, and the listener is thereby better able to perceive both temporal and nontemporal (e.g., pitch) features of the to-be-judged event. On this view, attention’s finer focus in time promotes selective attention to individual dimensions and results in better discrimination of both duration and pitch-change of discrete sequences relative to continuous glides. This view is supported by separate studies conducted by Henry (2007) and MacKenzie (2007) who observed that the magnitude of the auditory kappa effect (i.e., dependence of time judgments on perceived pitch) was attenuated when the target tone was preceded by an isochronous context sequence.

An alternative explanation for the observed differences between continuous and discrete stimuli in Part I comes from work by Zatorre and colleagues (Zatorre & Belin, 2001; Zatorre, Belin, & Penhune, 2002; Zatorre, Evans, Meyer, & Gjedde, 1992). Zatorre suggests that there is a trade-off between the temporal and spectral resolution achievable by the auditory system. Temporal and spectral processing are conceptualized in signal processing terms, where analysis along either dimension is constrained to be conducted within a limited-duration window. For tasks that require fine temporal resolution, a narrow temporal window is required. However, for
tasks that require fine spectral resolution, a longer window is desired; a longer window allows
the pitch processing system to integrate over more periods (i.e., cycles) of the sound, providing a
richer pitch representation. Temporal processing is assumed to be underpinned by left auditory
cortex, while spectral processing is supported by analogous right hemisphere locations. The
relevance of this theory with respect to the current studies is related to the possibility that the
isochronous timing of events for discrete stimuli promotes temporal processing, while
continuously changing frequency information in continuous stimuli promotes spectral
processing. On this view, differences in results for the two stimulus types may reflect operation
of different processing styles underpinned by lateralized neural mechanisms (Obleser, Eisner, &

Independent of the explanation for the observed differences between continuous and
discrete stimuli in Part I, a practical implication is suggested. The nature of the relation between
music and language is a topic that has recently begun to receive more attention in the area of
cognitive science (Patel, 2008). In this regard, it is worthwhile to note that music is one of the
few domains in which listeners experience discrete, and often isochronous sequences of events.
On the other hand, speakers tend to produce continuously changing patterns of acoustic
stimulation. This difference is reflected in the stimuli used to study music and speech in the
laboratory. I would like to suggest the possibility that differences between music and speech
observed experimentally should be interpreted cautiously, as these differences may reflect basic
differences between processing of continuous versus discrete auditory stimuli.

*Parts I and II: Continuous versus discrete stimuli*

A puzzle emerges when the results of the Part I and Part II experiments are compared. In
Part I, large differences between continuous and discrete stimuli were observed; listeners were
able to attend selectively to time and pitch when stimuli were discrete sequences, but not when stimuli were continuous tone glides. However, in Part II, results were very similar for both continuous and discrete stimuli; listeners used joint frequency-time information conveyed by both continuous and discrete stimuli to extrapolate the cue trajectory to a future signal location. Two explanations for the difference between Part I and Part II results are suggested, one of a bottom-up nature, and the other of a more top-down nature.

The first possibility for the difference between Parts I and II concerns acoustic differences in stimuli presented in the two parts. Specifically, in Part II, all stimuli were presented against a background of broadband noise. If it is assumed that Part I differences were due, at least in part, to continuous and discrete stimuli co-opting spectral and temporal processing wetware, respectively (Obleser, et al., 2008; Zatorre & Belin, 2001; Zatorre, et al., 1992), then it becomes plausible that the presence of a great deal more frequency information in Part II, i.e., broadband noise, could have tipped stimulus processing towards a spectral analysis on a purely acoustic basis, leading to more similar processing of continuous and discrete stimuli in Part II than in Part I. This hypothesis is easily testable by 1) adding broadband spectral information to the Part I stimuli, and 2) removing broadband spectral information from Part II stimuli.

The second possibility is that differences between Parts I and II were driven by the nature of the task that listeners were required to perform. In Part I, listeners were instructed to attend selectively, judging either the duration or pitch-change of stimuli without reference to the irrelevant dimension. In Part II on the other hand, although instructions about strategy were not provided, the best performance was achievable by attending jointly to the frequency-time trajectory of stimuli; listeners may have picked up on this and formed a listening strategy to this
effect. This explanation suggests that the behavioral differences for continuous and discrete stimuli in Part I but not in Part II were due to differences in conscious strategies used by listeners that differed between tasks. Previous work provides mixed support for this suggestion. Jones et al. (2006) observed that the degree to which listeners tuned attention jointly in pitch and time in a pitch-judgment task depended on instructions to the listener to attend to the pitch dimension of stimuli. However, the results of Henry and McAuley (in press) indicated that even when listeners were instructed to judge velocity (that is, joint pitch and time information), they were unable to do so and made their judgments based only on timing information. In the present case, the use of a conscious strategy also seems unlikely given that most listeners expressed surprise when they were debriefed and informed that most signals occurred at the frequency-time location based on extrapolation of the stimulus trajectory.

The first, acoustic explanation implies a process that listeners may not have direct access to, while the second, instructions-based explanation suggests differences in strategies consciously adopted by the participants. Zatorre (2007) has previously proposed that the degree to which listeners rely on a left-hemisphere temporal processing mechanism or a right-hemisphere spectral processing mechanism depend on both acoustic and knowledge-based cues. Thus it is possible that both bottom-up and top-down processes are at play here. However, further investigation in this area will be informative about fundamentals of pitch and time processing in discrete versus auditory stimuli.

*Ascending versus descending stimuli*

In Part II of this dissertation, differences were observed between ascending and descending cues such that listeners tuned their attention at or near the expected signal location following ascending cues, but overshot the expected signal frequency following descending cues.
Moreover, there were large differences in observed sensitivity to detect the signal, especially for continuous tone glide cues; listeners were more sensitive to detect the signal following descending relative to ascending cues. A number of hypotheses for this asymmetry were suggested in Chapter 6, thus only a summary will be provided here; however, some additional space will be spent discussing the possibility of an embodied explanation for differences between ascending and descending stimuli.

One suggestion for ascending / descending cue differences was that listeners have internalized and applied knowledge of gravity to a listening situation (Henry & McAuley, 2009; Henry, et al., 2009). Previous research has suggested that observers’ memory for visual object location is shifted down relative to the object’s true location, in line with the operation of an internalized knowledge of gravity (Freyd, et al., 1988; Hubbard & Bharucha, 1988). Previous auditory research has also demonstrated asymmetries between ascending and descending stimuli that are consistent with application of an auditory gravity principle (Hubbard, 1995a). One criticism of this view is the seeming lack of adaptive significance for an internalization of gravity in the auditory domain. However, one possible mechanism is suggested by the close association of music with movement (Chen, Penhune, & Zatorre, 2008; Haga, 2008; Phillips-Silver & Trainor, 2007; Toiviainen, Luck, & Thompson, 2009). Eitan and Granot (2004a, 2004b) observed a strong association of downward movement of an imagined character with descending pitch intervals; however, there was not a strong association of upward movement with ascending pitch intervals. If the reason for an asymmetry between ascending and descending cues is related to the association of music with movement, it should perhaps not be surprising that listeners apply a gravity-like principle to sounds.

A second possibility for the ascending / descending cue differences concerns the ease or
speed with which the human voice produces ascending versus descending pitch trajectories. Previous research indicates that descending pitch intervals are produced faster than ascending pitch intervals by untrained singers (Sundberg, 1973; Xu & Sun, 2000, 2001). The suggestion is that an embodied knowledge of one’s own vocal capabilities influences perception of auditory stimuli; a descending pitch trajectory may be heard as faster than its ascending counterpart because of the faster possible production speed for the descending stimulus.

Both explanations are similar to an embodied cognition approach to auditory perception. In this regard, Schubotz and colleagues (Schubotz, 2007; Schubotz & Von Cramon, 2003; Schubotz, Von Cramon, & Lohmann, 2003) have suggested a role for the lateral premotor cortex in the simulation and prediction of auditory events that are characterized by a high degree of regularity along the time and pitch dimensions. Specifically, it is suggested that timing and pitch prediction depend on premotor areas involved in articulation. Thus, this supports the idea that the ascending / descending cue asymmetry observed here may be related to simulation of sound production by the human voice. Several hypotheses are suggested by the possibility that simulation of sound production carried out by premotor brain regions may underlie the differences observed between ascending and descending cues in Part II of this dissertation. For example, trained singers do not show pronounced differences between transition speed for ascending versus descending pitch excursions (Sundberg, 1973); thus, a comparison of individuals with and without voice training should reveal an asymmetry for ascending versus descending intervals for listeners without voice training that is not present in trained singers. A second finding from voice production work is that the pitch transition rate increases with increasing interval width (Sundberg, 1973; Xu & Sun, 2000, 2001). Thus, listeners should tend to overshoot the expected signal frequency more so when the total frequency change of the cue is
relatively large. Clearly, further research will be necessary to clarify the role of an embodied knowledge of production capabilities or environmental invariants on auditory perception.

**Implications for pitch-time interdependence**

The degree to which pitch and time are independent versus interdependent in perception has received much recent attention (Boltz, 1989b, 1993; Henry & McAuley, 2009; Henry, et al., 2009; Jones, et al., 2006; Jones & Yee, 1993; Monahan & Carterette, 1985; Palmer & Krumhansl, 1987; Peretz & Coltheart, 2003; Peretz & Morais, 1989). Overall, the data are mixed, with studies of amusic listeners and neuroimaging work providing evidence for perceptual independence (Hyde & Peretz, 2004; Peretz, et al., 2002; Peretz, et al., 2009; Peretz, Gagnon, Hébert, & Macoir, 2004) and behavioral and recent neurophysiological work providing support for perceptual interdependence (Boltz, 1989b; Boltz & Jones, 1986; Jones & Ralston, 1991; Jones, et al., 1987; Neuhaus & Knösche, 2008). To reconcile the seeming conflicts in the literature, a number of researchers have suggested that factors such as musical training (Monahan, et al., 1987; Monahan & Carterette, 1985; Neuhaus & Knösche, 2008; Pitt & Monahan, 1987), global versus local processing (Monahan, et al., 1987), joint accent structure (Boltz, 1998; Jones & Ralston, 1991; Jones, et al., 1982), task demands (Prince, Schmuckler, & Thompson, 2009), and tonality (Prince, Schmuckler, et al., 2009; Prince, Thompson, & Schmuckler, 2009) modulate the nature of the relationship of pitch and time in perception.

The auditory motion hypothesis frames varying degrees of perceptual interdependence in terms of the extent to which listeners pick up on and make use of motion or trajectory information in auditory stimuli (Henry & McAuley, 2009; Henry, et al., 2009; Jones & Yee, 1993; MacKenzie, 2007). For auditory stimuli that convey a strong sense of motion, the observed degree of perceptual interdependence is expected to be large. On the other hand, for stimuli that
do not convey a strong sense of motion, the expectation is that pitch and time will be more likely to be interpreted as independent. Thus, the auditory motion hypothesis stresses the clarity of trajectory information in auditory stimuli. In this regard, a potential avenue for future research would be to assess the degree to which factors like joint accent structure (Boltz, 1998; Jones & Ralston, 1991; Jones, et al., 1982), task demands (Prince, Schmuckler, et al., 2009), and tonality (Prince, Schmuckler, et al., 2009; Prince, Thompson, et al., 2009) affect clarity of or attention to motion information. Along these lines, one factor that has been previously shown to be important is velocity (Henry & McAuley, 2009; Henry, et al., 2009); increasing the velocity of discrete tone sequences increases reliance on trajectory information. In Part I, whether a stimulus moves continuously or implies motion in pitch space was shown to be important with respect to the clarity of motion information conveyed by the stimulus. The results of Part II indicate that discrete stimuli can convey trajectory information as continuous stimuli, but the factors that make this possible remain to be more fully elaborated.

In Part I of this dissertation, an imputed pitch velocity model was used to compute the magnitude of pitch-time interdependence. This model was previously applied to the auditory kappa and tau effects (Henry & McAuley, 2009; Henry, et al., 2009). The model yields a dependent variable, \( w \), which quantifies the magnitude of perceptual interdependence when fitted to response proportion data. The value \( w \) possesses the positive characteristic that its scale is not tied to the specific task; values of \( w \) are comparable across tasks and participants, and so provide a metric for evaluating the influence of stimulus and task factors on the magnitude of pitch-time interdependence. Going forward, application of the imputed pitch velocity model in combination with neuroimaging work using both fMRI and EEG technologies will help to elaborate the stimulus / task features that influence the degree of pitch-time interdependence as well as the
processing stage(s) at which pitch and time information are combined.

*Implications for attention and auditory scene analysis*

I have introduced the tenets of dynamic attending theory earlier in this dissertation; however, I will provide a brief review here. Dynamic attending theory (Jones, 1976, 2004; McAuley & Jones, 2003) is built on the observation that many naturally-occurring events, including music and speech, are inherently rhythmic; dynamic attending theory also proposes that attention is rhythmic. Through entrainment, listeners’ attentional rhythms become synchronized with the rhythms of external stimulating events; attentional peaks are assumed to happen coincident with external event onsets. Moreover, attentional rhythms are predictive and self-sustaining. Thus, once coupled with an external rhythm, peaks in the attentional rhythm predict the timing of future event onsets, constituting the generation of expectations by the listener. Expectations are satisfied when events happen at an expected time, and expectations are violated when events occur early or late with respect to peaks of the attentional rhythm.

Attentional rhythms are suggested to be extended in pitch space; however, how this is accomplished is less well articulated than the mechanism for temporal rhythms, which has been formally defined in several models (Large & Jones, 1999; McAuley, 1995) and tested with respect to quantitative predictions (McAuley & Jones, 2003). Jones (1976) first proposed the idea that rhythmic attentional energy is distributed in a two-dimensional pitch-time plane, and the current results support this hypothesis. Recently, Jones et al. (2006) have articulated a pitch-time entrainment hypothesis, and provided a test of derived hypotheses with mixed results. Specifically, listeners judged the relative pitch of a target tone embedded in an isochronous sequence that was designed to tune listeners’ attention in pitch either grossly or more narrowly. Pitch judgments were found to be more accurate when attention was tuned narrowly in pitch and
to the correct time of the target tone. However, no interactions between pitch and time tuning were observed, inconsistent with the idea that listeners made use of a two-dimensional pitch-time tuning mechanism. It should be noted that in this study, patterns were periodic (i.e., alternating between high and low relative pitch values), and so the extent to which attention was tuned along a trajectory per se is questionable. However, the results of the current dissertation strongly support the hypothesis that listeners are sensitive to the pitch-time trajectory of unfolding auditory stimuli.

A different theoretical perspective that also addresses the possibility that listeners make use of pitch-time trajectory information is that of Bregman and colleagues in the realm of auditory scene analysis; for a review see Bregman (1990). This account is concerned with the cues that listeners use to organize their auditory environments into events, or streams, each of which is attributable to a single sound source. Bregman’s ideas about the cues listeners use to accomplish this analysis are rooted deeply in the Gestalt principles that describe organization of the visual environment into figure and ground. In particular, principles of similarity (timbre) and proximity (pitch) are proposed to be especially important for auditory scene analysis.

In a number of studies, Bregman and his colleagues have entertained the possibility of a trajectory-based grouping principle analogous to the law of good continuation (Ciocca & Bregman, 1987; McPherson, Ciocca, & Bregman, 1994; Steiger & Bregman, 1981; Tougas & Bregman, 1985, 1990). In these studies, trajectory-based groupings are pitted against proximity-based groupings, and the failure to observe dominance of the trajectory-based grouping has been interpreted as evidence that listeners do not use pitch-time trajectories of sounds when forming auditory streams (Ciocca & Bregman, 1987; Dannenbring, 1976; Kluender & Jenison, 1992; Tougas & Bregman, 1985, 1990).
Bregman, 1981; Tougas & Bregman, 1985), listeners were presented with auditory stimuli that could have been interpreted as ‘bouncing’ or ‘crossing’ depending on whether the listener made use of a pitch proximity or trajectory principle for grouping, respectively (see Figure 34A). In these studies, listeners first heard a standard stimulus that clearly conveyed either a bouncing or crossing percept, then rated the clarity of the presence of the standard percept in the ambiguous comparison. For both continuous and discrete stimuli, bouncing was reported as being the clearer percept unless a secondary cue (e.g., timbre) suggested a crossing percept. These results are consistent with reliance on a pitch proximity principle rather than a trajectory-based grouping principle.

However, a study by McPherson, Ciocca, and Bregman (1994), consistent with the results reported here, suggests that listeners make use of velocity information to anticipate the future time course of auditory events, and assign temporally extended patterns of auditory stimulation to a single auditory stream. In this study, listeners indicated the clarity of the presence of either a bouncing or crossing percept, as described above. The novel aspect of this study was that the crossing glide segments had different velocities; a stimulus diagram is given in Figure 34B. Velocity was manipulated by holding pitch-change constant and varying duration (Exp. 1, 2, 3) and by holding duration constant and varying pitch-change (Exp. 4). Listeners demonstrated a bias to hear the glides as crossing, which was especially pronounced when constant-pitch segments connecting glide pairs was removed and crossing / bouncing glides were instead separated by a period of silence. This study also demonstrated that listeners heard the glides as crossing, even when the point of intersection was replaced by silence, indicating that the crossing percept was supported by an extrapolation process that did not depend on hearing the glides as
Figure 34. (A) A schematic of stimuli from studies by Tougas and Bregman (1985, 1990). The percept could be either of two glides that crossed, or a high and a low sound that bounced when they met in the middle of the pitch range. Listeners were given a prototype crossing or bouncing standard, and asked to rate the clarity of the presence of the standard in the stimulus. (B) A schematic from the study by McPherson et al. (1994). The task was the same, but the velocity of the crossing glide segments was different. This manipulation increased the likelihood of hearing the crossing, as opposed to the bouncing, percept.
Auditory Motion for Continuous and Discrete Sounds

continuous. This study is regarded as particularly important, because although glide slope (i.e., velocity) has been treated as an independent variable in several previous studies (Ciocca & Bregman, 1987; Kluender & Jenison, 1992), only the McPherson et al. (1994) study allows for the possibility that velocity guides the parsing of the auditory environment into streams, that is, auditory scene analysis. Taken together with McPherson’s results, the results of the experiments presented here are consistent with the auditory motion hypothesis, which assumes that velocity (i.e., the rate of pitch change over time) plays a particularly important role in the guidance of auditory attention and in auditory perception more generally. Because this issue has received relatively little attention, it is suggested that future research should focus on clarifying the cognitive and neuropsychological mechanisms by which listeners make use of velocity information, the stage in processing at which velocity information is important, and the stimulus and task characteristics that promote use of velocity information.

Conclusions

The aim of this dissertation was to address the possibility that listeners make use of velocity information to generate expectations about the future time course of continuous and discrete auditory stimuli. Toward this end, two sets of experiments were conducted. In Part I, listeners judged the time-change or pitch-change of a variable-velocity comparison stimulus relative to a constant-velocity standard stimulus; stimuli were either continuous tone glides or discrete tone sequences. Consistent with the auditory motion hypothesis, listeners’ judgments about duration and pitch-change were strongly influenced by velocity when stimuli changed continuously in pitch space; however, no perceptual distortions were observed for discrete tone sequences, suggesting that motion information was less clearly conveyed or easier to ignore than for tone glides. In Part II, listeners detected signals that occurred at expected or unexpected
frequency-time locations based on extrapolation of the trajectory of continuous or discrete auditory cues through a variable duration noise burst in order to test the hypothesis that attention is tuned jointly by frequency-time velocity information. Again consistent with the auditory motion hypothesis, for both continuous and discrete stimuli, listeners used velocity information to tune attention to the frequency-time location of the signal; moreover, large asymmetries were observed for descending relative to ascending trajectories, with listeners overshooting the location of the expected signal following descending cues. Taken together, the current results support the auditory motion hypothesis, indicating that listeners make use of velocity information conveyed by auditory stimuli to generate expectations about the pitch-time locations of future events so that they may better perceive an information-carrying event that fulfills this expectation.
REFERENCES


Auditory Motion for Continuous and Discrete Sounds


Auditory Motion for Continuous and Discrete Sounds


Key Terms

Pitch, Pitch Contour, and Pitch Interval

Pitch refers to the psychological correlate of frequency, and describes how low or high an auditory stimulus sounds. A melody is a serial pattern of pitches, where a melody’s identity is determined by both its pitch contour and specific pitch intervals. Pitch contour refers to the overall pattern of ups and downs in a melody, and pitch interval refers to the specific distance in semitones (STs), or musical half-steps, between pairs of notes. In Western music, certain pitch intervals and patterns of pitch intervals are more expected than others, and are determined by the key in which a piece is played (Krumhansl, 2001); the hierarchy of expected pitch relations is referred to as tonality.

Rhythm

Here, the term rhythm refers to a serial pattern of durations, although it should be noted that the term is used in other work to describe the perception of a durational pattern, which may differ from the durational pattern present in the signal (Handel, 1989). In music, rhythm is conveyed by serially ordering notes with different durations (e.g., quarter notes, half notes) and rests (silences).

Beat and Meter

Beat refers to a regular, periodic ‘pulse’ in music (Large & Palmer, 2002; Parncutt, 1994). It is the beat that listeners spontaneously tap their feet or bob their heads to when they move to music. Meter is a hierarchical organization of musical time, and defines the number of beats that occur in a measure. Thus the meter determines the pattern of strong and weak beats in a piece, that is, whether the piece sounds like, for example, a march (ONE-two-ONE-two) or a
waltz (ONE-two-three-ONE-two-three).

**Accentuation and Accent Structure**

The examples of the march and waltz above highlight an important property of music, which is that not all events are heard as equally strong. Both pitch and time properties of music contribute to which events are heard as accented and to the relative salience of accented events, that is, accent structure. With respect to the pitch dimension, accents coincide with changes in pitch direction, i.e., contour changes, and locations of relatively large pitch intervals. Pitch accents can also be marked by the presence of a tonally stable note, that is, a note that is more expected based on the key in which the piece is played. With respect to the time dimension, accents occur at the location of relatively long-duration events and following silences (rests). Generally, both pitch accents and time accents occur periodically, that is, at regular temporal intervals. Moreover, pitch and time accents typically coincide. Music in which the relationship of periodic pitch and time accents forms a simple ratio (e.g., 1:1, 2:1) are termed concordant, while accent relationships that form complex ratios (e.g., 3:2) are termed discordant.

**Montreal Battery of Evaluation of Amusia (MBEA)**

The MBEA (Peretz, Champod, & Hyde, 2003) is the standard battery for assessment and diagnosis of amusia (Cuddy, Balkwill, Peretz, & Holden, 2005; Douglas & Bilkey, 2007; Foxton, et al., 2004; Patel, Iversen, Chen, & Repp, 2005; Sloboda, Wise, & Peretz, 2005). The MBEA consists of six subtests: Scale, Contour, Interval, Rhythm, Meter, and Memory. Four of the subtests (Scale, Contour, Interval, Rhythm) present pairs of melodies and ask listeners to make a *same/different* judgment. The remaining two subtests (Meter, Memory) involve presenting a single melody; listeners make a *march/waltz* judgment for the Meter subtest and an *old/new* judgment for the Memory subtest. The Scale, Contour, and Interval subtests are
presumed to be diagnostic of melodic pitch perception abilities, whereas the Rhythm and Meter subtests are proposed to be diagnostic of temporal processing capabilities. With respect to diagnosis, a listener is typically classified as tone-deaf if his/her aggregate proportion correct (PC) score across the six subtests falls more than two standard deviations below the mean of a normative sample (Peretz, et al., 2003).

Methods

Magnetic Resonance Imaging (MRI)

MRI is a non-invasive imaging technique that takes advantage of the magnetic properties of nuclei in the brain to produce a high-resolution anatomical image. Specifically, a series of magnetic gradients is quickly switched on and off, ‘tipping’ hydrogen nuclei relative to the scanner’s main magnetic field. Images are then constructed on the basis of differences in the degree of relaxation back to the main magnetic field at the time of measurement, which differs between tissue types (i.e., white matter versus gray matter).

Voxel-Based Morphometry. Voxel-based morphometry is a technique for comparing brain structure between participants (Ashburner & Friston, 2000). Specifically, the technique involves comparing the gray matter volume between participants on a voxel-by-voxel basis, where a voxel, similar to a pixel, refers to a location in three-dimensional brain space. The technique is used to compare cortical volume in regions of interest between individuals or groups of individuals.

Diffusion Tensor Imaging. Diffusion tensor imaging is based on the properties of diffusion of molecules, for example water, within and outside of white matter tracts in the brain (Le Bihan, et al., 2001). Specifically, diffusion refers to the random motion of molecules unconstrained in space (i.e., Brownian motion). Diffusion takes place more quickly in the
direction of axonal fibers than in the perpendicular direction; thus diffusion rate is informative regarding fiber tract direction. This technique is informative with respect to connectivity (i.e., white matter) between regions of interest in individuals.

**Functional MRI.** Functional MRI (fMRI) is an imaging technique used to isolate brain regions that are involved in processing, that is, brain regions that are ‘active’ during a task (Matthews & Jezzard, 2004; Ramsey, Hoogduin, & Jansma, 2002). During processing, an active part of the brain uses oxygen, increasing the level of deoxygenated blood in the region. The brain’s response is to send oxygenated blood to the region to replace the depleted oxygen; this is referred to as the hemodynamic response. Deoxygenated and oxygenated blood have different magnetic properties, and it is this contrast that is of interest in functional images. The blood oxygenated level dependent (BOLD) response during a task is taken as an index of the involvement of a brain region in a cognitive task. FMRI provides fine spatial resolution, on the order of 3 mm. However, because hemodynamic response is a relatively slow process, peaking around 6-8 seconds after a to-be-processed stimulus is presented, temporal resolution with this technique is relatively poor.

**Positron Emission Tomography (PET)**

PET scanning involves injecting a radioactive tracer into the participant, which emits positrons, which in turn emit photons at the time of a collision with an electron; it is the emission of photons resulting from annihilation that are measured in PET. To detect brain activation during a cognitive task, the logic underlying PET is similar to fMRI. In particular, the radioactive tracer is bound to molecules, often oxygen. When the brain sends oxygenated blood to an active region, a relatively high concentration of the tracer will be detected in this area. The spatial and temporal resolution of PET is poorer than that of fMRI, but PET scanning is more sensitive in
terms of signal-to-noise ratio and can provide information about relative concentrations of neurotransmitters in brain regions of interest (Volkow, Rosen, & Farde, 1997).

Electroencephalography (EEG)

EEG involves monitoring modulations of electrical activity on the scalp that result from neuronal firing in the cortex (Janata, 2001; Volkow, et al., 1997). Because the activity is detected on the scalp, EEG has inherently poor spatial resolution. However, the temporal resolution is far superior to fMRI or PET, and is on the order of milliseconds. Modulations in electrical potentials that are time-locked to stimulus onsets are referred to as event-related potentials (ERPs), and ERPs are analyzed in terms of components that are labeled according to their polarization and time of occurrence. For example, a transient negativity occurring 100 ms after stimulus onset would be an N100, or N1.
APPENDIX B: Research Participant Information and Consent Form

You are being asked to participate in a research project. Researchers are required to provide a consent form to inform you about the study, to convey that participation is voluntary, to explain risks and benefits of participation, and to empower you to make an informed decision. You should feel free to ask the researchers any questions you may have.

Study Title: Pitch and Time Perception
Researcher and Title: Dr. Devin McAuley, Associate Professor; Molly. J. Henry, M. A., Visiting Scholar
Department and Institution: Department of Psychology, Michigan State University
Address and Contact Information:

Dr. Devin McAuley
282B Psychology Building
Michigan State University, East Lansing, MI 48824
Email: mcauleyj@msu.edu
Phone: 517-353-9069

Molly J. Henry
55 Psychology Building
Michigan State University, East Lansing, MI 48824
Email: henrymol@msu.edu
Phone: 517-353-5022

1. PURPOSE OF RESEARCH:

You are being asked to participate in a research study aimed at learning more about how people process pitch and time when they listen to events in their environment. You have been selected as a possible participant in this study because you expressed an interest in participating and are a member of the MSU community who is at least 18 years of age with self-reported normal hearing.

From this study, we hope to learn more about how people process pitch and time when they listen to music, speech, and other natural environmental events. Understanding the ways in which people pitch and time helps us to know more about how we perform tasks that seem simple to the listener yet are complicated scientific questions. Examples include recognizing the voice of a friend, coordinating with a group to make music, coordinating with music to produce dance, and taking turns in a conversation.

In total, about 500 people are being asked to participate in the study with the time of a session varying between 15 minutes and 3 hours. Different people will be asked to attend between 1 and 3 sessions. You are being asked to attend _____ sessions, each lasting approximately _____ minutes of your time.

2. WHAT YOU WILL DO:

If you agree to participate, then the following may happen as part of the research study. You may be asked to participate in up to three (specified above) session(s) in which you will listen to sounds and make simple judgments about the pitch (e.g., ‘high’ / ‘low’), timing (e.g., ‘short’ / ‘long’), or meter (‘march’ / ‘waltz’) of these sounds. Sounds may be single tones, short sequences made up of several tones, or short melodies. Your response will be made either by providing a written response or by pushing buttons and/or typing on a special response device or computer mouse/keyboard. Any sounds that you will hear will be presented at a comfortable listening level.

In addition to making simple judgments about sounds, you will be asked to complete two surveys. One survey is a background survey that includes questions about age, gender, musical training, linguistic background, and any over-
the-counter or prescription medications that you took prior to participating in the study. The second survey addresses your impressions of the tasks you will perform in the study.

The research findings are likely to be published in peer-reviewed research journals, and these are available to the public. The results of the research study will not be provided to individual participants.

3. POTENTIAL BENEFITS:

There are no direct benefits to you for participating in this study. However, your participation in this study may contribute to improved understanding of how humans listen to and understand events in their environment. This will contribute to basic science research on mechanisms of pitch and time processing and may in turn help to direct future research on various neurological disorders that are associated with pitch and time processing problems that affect music and speech understanding, as well as perception of environmental events.

4. POTENTIAL RISKS:

There are no foreseeable risks associated with participation in this study. Any sounds you may hear will be presented at a comfortable listening level. The overall risks in this study are no greater than encountered in daily life.

5. PRIVACY AND CONFIDENTIALITY:

The data for this project will be kept confidential to the maximum extent allowable by law. All information provided is confidential and all paper records of your research participation will not include your name and will be stored in a secure filing cabinet. All responses you make on the computer and associated data files will be maintained on a password protected computer or secure backup storage device in the primary researcher’s MSU office for at least three years after the project closes and will not be linked to your name. The results of this study may be published or presented at professional meetings, but the identities of all research participants will remain confidential. Data will be stored indefinitely. Only the researchers and the IRB will have access to the data.

6. YOUR RIGHTS TO PARTICIPATE, SAY NO, OR WITHDRAW:

Participation in this research project is completely voluntary. You have the right to say no. Choosing not to participate or withdrawing from this study will not make any difference in benefits to which you are otherwise entitled. You may change your mind at any time and withdraw from the study. If you withdraw from the study before the experiment is completely over, you will be compensated with course credit or monetarily (as indicated below) for the amount of time you spent participating in the study. You may choose not to answer specific questions or to stop participating at any time. Whether you choose to participate or not will not affect your grade or any relationship you have with Michigan State University.

7. COSTS AND COMPENSATION FOR BEING IN THE STUDY:

For participating in this study, you will receive only one of the following forms of compensation:

- [ ] Course credit commensurate with time spend during study participation.
☐ Monetary compensation at a rate of $10/hr.

If monetary compensation is selected, payment for participation will take place within 3 weeks.

8. CONTACT INFORMATION FOR QUESTIONS AND CONCERNS

If you have any concerns or questions about this study, such as scientific issues, how to do any part of it, or to report an injury, please contact the researcher (Molly J. Henry, 55 Psychology Building, Michigan State University, henrymol@msu.edu, 517-353-5022, or her advisor Dr. Devin McAuley, 282B Psychology Building, Michigan State University, mcauleyj@msu.edu, 517-353-9069).

If you have questions or concerns about your role and rights as a research participant, would like to obtain information or offer input, or would like to register a complaint about this study, you may contact, anonymously if you wish, the Michigan State University’s Human Research Protection Program at 517-355-2180, Fax 517-432-4503, or email irb@msu.edu or regular mail at 207 Olds Hall, MSU, East Lansing, MI 48824.

9. DOCUMENTATION OF INFORMED CONSENT:

Your signature below means that you voluntarily agree to participate in this research study.

_________________________________________   _____________________________
Signature       Date

You will be given a copy of this form to keep.
APPENDIX C: Research Participant Information and Consent Form

You are being asked to participate in a research project. Researchers are required to provide a consent form to inform you about the study, to convey that participation is voluntary, to explain risks and benefits of participation, and to empower you to make an informed decision. You should feel free to ask the researchers any questions you may have.

Study Title: Attention in Frequency and Time
Researcher and Title: Dr. Devin McAuley, Associate Professor; Molly. J. Henry, M. A., Visiting Scholar
Department and Institution: Department of Psychology, Michigan State University
Address and Contact Information:

Dr. Devin McAuley  
282B Psychology Building  
Michigan State University, East Lansing, MI 48824  
Email: mcauleyj@msu.edu  
Phone: 517-353-9069

Molly J. Henry  
55 Psychology Building  
Michigan State University, East Lansing, MI 48824  
Email: henrymol@msu.edu  
Phone: 517-353-5022

1. PURPOSE OF RESEARCH:

You are being asked to participate in a research study aimed at learning more about how people pay attention to frequency (pitch) and time when they listen to music and speech. You have been selected as a possible participant in this study because you expressed an interest in participating and are a member of the MSU community who is at least 18 years of age with self-reported normal hearing.

From this study, we hope to learn more about how people pay attention to pitch and time simultaneously when they listen. Understanding the interactions of pitch and time is important to better understand how people listen to music and speech in noisy environments, and may contribute to the limited knowledge we have of a disorder called amusia (tone-deafness).

In total, about 250 people are being asked to participate in the study with the time of a session varying between 15 minutes and 3 hours. Different people will be asked to attend between 1 and 3 sessions. You are being asked to attend _____ sessions, each lasting approximately _____ minutes of your time.

2. WHAT YOU WILL DO:

If you agree to participate, then the following may happen as part of the research study. You may be asked to participate in session in which you will listen to a series of sounds and make simple judgments about whether you hear a quiet ‘beep’ at the end, either by providing a written response, by pushing buttons and/or typing on a special response device or computer mouse/keyboard. Any sounds that you will hear will be presented at a comfortable listening level.

In addition to judging whether you hear the beep following the sounds you hear, you will be asked to complete two surveys. One survey is a background survey that includes questions about age, gender, musical training, linguistic background, and any over-the-counter or prescription medications that you took prior to participating in the study. The second survey addresses your impressions of the tasks you will perform in the study.
The research findings are likely to be published in peer-reviewed research journals, and these are available to the public. The results of the research study will not be provided to individual participants.

3. POTENTIAL BENEFITS:

There are no direct benefits to you for participating in this study. However, your participation in this study may contribute to improved understanding of how humans listen to and understand events in their environment. This will contribute to basic science research on mechanisms of frequency and time processing and may in turn help to direct future research on various neurological disorders that are associated with frequency and timing difficulties.

4. POTENTIAL RISKS:

There are no foreseeable risks associated with participation in this study. Any sounds you may hear will be presented at a comfortable listening level. The overall risks in this study are no greater than encountered in daily life.

5. PRIVACY AND CONFIDENTIALITY:

The data for this project will be kept confidential to the maximum extent allowable by law. All information provided is confidential and all paper records of your research participation will not include your name and will be stored in a secure filing cabinet. All responses you make on the computer and associated data files will be maintained on a password protected computer or secure backup storage device in the primary researcher’s MSU office for at least three years after the project closes and will not be linked to your name. The results of this study may be published or presented at professional meetings, but the identities of all research participants will remain confidential. Data will be stored indefinitely. Only the researchers and the IRB will have access to the data.

6. YOUR RIGHTS TO PARTICIPATE, SAY NO, OR WITHDRAW:

Participation in this research project is completely voluntary. You have the right to say no. Choosing not to participate or withdrawing from this study will not make any difference in benefits to which you are otherwise entitled. You may change your mind at any time and withdraw from the study. If you withdraw from the study before the experiment is completely over, you will be compensated with course credit or monetarily (as indicated below) for the amount of time you spent participating in the study. You may choose not to answer specific questions or to stop participating at any time. Whether you choose to participate or not will not affect your grade or any relationship you have with Michigan State University.

7. COSTS AND COMPENSATION FOR BEING IN THE STUDY:

For participating in this study, you will receive only one of the following forms of compensation:

☐ Course credit commensurate with time spend during study participation.

☐ Monetary compensation at a rate of $10/hr.

If monetary compensation is selected, payment for participation will take place within 3 weeks.
8. CONTACT INFORMATION FOR QUESTIONS AND CONCERNS

If you have any concerns or questions about this study, such as scientific issues, how to do any part of it, or to report an injury, please contact the researcher (Molly J. Henry, 55 Psychology Building, Michigan State University, henrymol@msu.edu, 517-353-5022, or her advisor Dr. Devin McAuley, 282B Psychology Building, Michigan State University, mcauleyi@msu.edu, 517-353-9069).

If you have questions or concerns about your role and rights as a research participant, would like to obtain information or offer input, or would like to register a complaint about this study, you may contact, anonymously if you wish, the Michigan State University’s Human Research Protection Program at 517-355-2180, Fax 517-432-4503, or email irb@msu.edu or regular mail at 207 Olds Hall, MSU, East Lansing, MI 48824.

9. DOCUMENTATION OF INFORMED CONSENT:

Your signature below means that you voluntarily agree to participate in this research study.

_________________________________________   _____________________________
Signature       Date

You will be given a copy of this form to keep.
APPENDIX D: CALIBRATING SIGNAL DETECTABILITY

For Experiments 8 – 10, it was important to calibrate the signal amplitude such that all listeners were performing at a level that corresponded to $d' = 1.5$. The method was similar to that previously used by Crum and Hafter (2008) and Green, McKey, and Licklider (Green, McKey, & Licklider, 1959). All listeners who participated in Experiments 8 – 10 detected pure-tone signals in noise in a paradigm similar to Experiment 7, except that signal intensity was varied parametrically, and stimulus presentation was according to the method of constant of stimuli (Urban, 1910). Psychometric curves were estimated for each listener as a function of signal intensity, and the intensity level corresponding to $d' = 1.5$ was estimated for each listener.

Signals were present on 50% of trials, and signal intensity took on values of 0 dB, 25 dB, 31 dB, 34 dB, 37 dB, 39 dB, 40 dB, 42 dB, 43 dB. On each trial, the frequency of the signal was drawn on each trial from a uniform distribution ranging between 850 Hz and 2150 Hz. Listeners detected pure-tone signals with randomly chosen frequencies in noise by responding yes (signal present) or no (signal absent). Listeners completed 54 trials in each of 10 blocks, responding once to each combination of trial type, noise burst duration, and signal intensity. In total, listeners completed 540 trials; 10 observations were collected for each combination of noise burst duration and signal intensity.

Proportions of correct responses for signal trials were determined as a function of signal intensity for each listener, averaged over noise burst duration. For each listener, a sigmoid function given by the equation

$$f(x) = \frac{1}{1+e^{-\gamma(x+\theta)}}$$

was fit to proportion correct data. In this equation, $\gamma$ provides a measure of the location on the x-axis corresponding the signal intensity judged to be present in the stimulus 50% of the time ($\gamma = -$
PSE), and $\theta$ provides an estimate of the slope of the psychometric function ($\theta = 1/JND$).

Target signal detectability was chosen to be $d' = 1.5$. Assuming zero bias, $d' = 1.5$ corresponds to $PC = 0.77$. Thus for each listener, equation 6 was solved for the $x$-value (signal intensity) that yielded $f(x) = 0.77$. On average, listeners detected the signal 77% of the time for a 36.4 dB signal ($SD = 0.8$ dB).
NOTES

1. One cent is 1/100 of a semitone, or musical half-step, evenly divided in log frequency units.

2. See Appendix B for the consent document used to obtain informed consent for the experiments conducted in Part I.

3. One participant did not report gender.

4. In Experiments 4 – 6, total duration of the standard tone sequence was 750 ms, which differs from the value of 1000 ms used in Experiments 1 – 3. Moreover, standard and comparison velocities were somewhat higher in Experiments 4 – 6 than in Experiments 1 – 3 (i.e., 667 Hz/s, 1333 Hz/s, 2000 Hz/s). These confounds were unintentional, and resulted from a technical error during stimulus generation. The possibility that these differences drove differences between the two sets of experiments is considered in the General Discussion of Chapter 3.

5. Historically, two theoretical perspectives can be contrasted with respect to attentional tuning to frequency (Swets, 1963). The single-band model proposed by Tanner (1956) suggested that listeners are capable of monitoring only one frequency band at one time. Thus, heightened sensitivity to detect to signals of different frequencies is accomplished by sweeping the listening band back and forth between the target frequencies. This model is conceptually similar to the search-light model of visual attention, in that movement of the filter is accomplished through analog sweeping of the intermediate frequency range. The multi-band model of Green (1958) assumed that all task-relevant frequency bands were monitored simultaneously, and the outputs of each band were summed before processing. Both views predict a decrement in performance as a function of the number of to-be-monitored listening bands, but neither quantitatively accounted for the magnitude of the
Auditory Motion for Continuous and Discrete Sounds

... decrement resulting from uncertainty. More recently, a consensus has been reached that both listening strategies are utilized, but individuals differ in the degree to which they use one or the other strategy (Swets, 1963), and different experimental situations promote the use of different listening strategies (Hübner & Hafter, 1995).

6. In the Egan et al. (1961) study, signals were presented at threshold intensities and were present only on 50% of trials; thus failing to detect a signal early in the interval did necessarily mean an increased likelihood of a signal presentation later in the interval, as in studies of foreperiod effects.

7. See Appendix C for the consent document used to obtain informed consent for the experiments conducted in Part II.

8. One listener did not have a hearing screening, but self reported normal hearing.

9. WM values reported in the text are means over 100 iterations of the Monte Carlo simulation.

10. I would like to thank Kathrin Lange for bringing this possibility to my attention.