Title
Processing Stimuli over Time: Musical Modes and Audiovisual Binding

Author
Mann, Daniel

Publication Date
2014

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,
IRVINE

Processing Stimuli over Time: Musical Modes and Audiovisual Binding

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Psychology

by

Daniel Mann

Dissertation Committee:
Professor Charles Chubb, Chair
Associate Professor Charles Wright
Professor Kourosh Saberi

2014

DEDICATION

To

my family, friends, colleagues, and mentors

in recognition of their worth

I couldn’t have done it without you.

TABLE OF CONTENTS

	Page
LIST OF FIGURES	iv
LIST OF TABLES	vi
ACKNOWLEDGMENTS	vii
CURRICULUM VITAE	viii
ABSTRACT OF THE DISSERTATION	xi
CHAPTER 1: The ingredients of major/minor mode discrimination: pitch-height sensitivity, detecting a scale violated melody, and more.	1
CHAPTER 2: End on a high note: Resolution changes the influence of pitch when discriminating major and minor modes.	28
CHAPTER 3: Binding brightness and loudness: the limits of attention and the temporal window	68
REFERENCES	99
APPENDIX A: Chapter 2 Stimuli: Note-Count Vectors	104
APPENDIX B: Markov Chain Monte Carlo Sampling Procedure	107

LIST OF FIGURES

		Page
Figure 1.1	Major/Minor Task Replication	13
Figure 1.2	Task Performance Grouped by Major/Minor Sensitivity	14
Figure 1.3	SVMC and PHC Performance Separated by MM Group	18
Figure 1.4	PM and PHC Performance Separated by MM Group	19
Figure 1.5	PM and PHC Performance Separated by Tonal Language	21
Figure 1.6	MM and SVMC Performance Separated by Tonal Language	22
Figure 2.1	Rhythmic Notation for Conditions	36
Figure 2.2	Target Function Example	37
Figure 2.3	Full Model Example	40
Figure 2.4	Extended-Notes Condition: Tone-Weighting Functions	47
Figure 2.5	Extended-Notes Condition: Temporal-Weighting Functions and Criterion Parameters	52
Figure 2.6	Rests Condition: Tone-Weighting Functions	54
Figure 2.7	Rests Condition: Temporal-Weighting Functions and Criterion Parameters	58
Figure 2.8	Non-Rhythmic Condition: Tone-Weighting Functions	59
Figure 2.9	Non-Rhythmic Condition: Temporal-Weighting Functions and Criterion Parameters	63
Figure 3.1	Stimulus Presentation	75
Figure 3.2	Target Filter Example	77
Figure 3.3	Target Filters for Each Condition	78
Figure 3.4	Temporal Binding Parameters	82

Figure 3.5	Full Model Results for Single-Modality Attention Conditions	88
Figure 3.6	Full Model Results for Binding Attention Conditions	89
Figure 3.7	Full Model Temporal Binding Parameters	91
Figure 3.8	Hierarchical Model Basis Functions	92
Figure 3.9	Hierarchical Model Alpha Parameters	93
Figure 3.10	Hierarchical Model Token-Weighting Functions	94
Figure 3.11	Hierarchical Model Temporal Binding Parameters	95

LIST OF TABLES

		Page
Table 1.1	Predictor Variables in Major/Minor Group Assignment	24
Table 2.1	Full Model Constraints and Degrees of Freedom for Extended-Notes and Rests Conditions	41
Table 2.2	Full Model Constraints and Degrees of Freedom for the Non-Rhythmic Condition	43
Table 2.3	Likelihood-Ratio Tests for Separability of Rhythm and Pitch for the Non-Rhythmic Condition	44
Table 3.1	Full Model Constraints and Degrees of Freedom	84
Table 3.2	Hierarchical Model Constraints and Degrees of Freedom	86
Table 3.3	Full Model Consistency Parameter Estimates	90

ACKNOWLEDGMENTS

I would like to thank my committee chair, Professor Charles Chubb. Over the last 5 years, he was a master of flexibility and mentorship. Charlie was very available when I needed him, but he also gave the space necessary for my growth as an independent researcher. The respect he showed to me and all of his students provided an especially comfortable learning environment, and his example laid the foundations for my personal teaching and mentorship philosophies.

I'm deeply grateful for my committee members, Associate Professor Charles Wright and Professor Kourosh Saberi. Charles (Ted) advised me for my undergraduate honors thesis- my first real research project. This particular experience and exposure to experimental psychology motivated my application for graduate school and pursuit of research. Kourosh advised me as soon as I entered the program, and his mentorship helped build my confidence as a new researcher. Both Ted and Kourosh continually supported me over the last 5-6 years and I'm so thankful.

The mentorship and support from each of my committee members extended far beyond the typical expectations for advisors.

In addition, I thank Professor Michael Lee for being encouraging and helpful over the last 6 years (before and throughout my Ph.D. program).

Financial support was provided by the University of California, Irvine and NSF Grant BCS-0843897.

CURRICULUM VITAE

Daniel Mann

EDUCATION

University of California, Irvine, CA
Ph.D. in Psychology **2014**
Concentration: Cognitive Neuroscience
Advisor Charles Chubb
Completed course sequences in quantitative and computational research methods that covered statistics, experimental design, programming for experiments and data analysis. Completed fMRI course series which included design, implementation, and data analysis of an original fMRI experiment. Other coursework includes Computational Bayesian Statistics, Bayesian Graphical Models, Fourier Analysis, Computational Neuroscience and the "Mind-Body problem."

B.A. Honors in Psychology **2009**
Minor: Linguistics
Honors Thesis: "Effector Selection and Violations of the Stimulus-Response Uncertainty Effect for Visually Guided Movements"

FELLOWSHIPS & AWARDS

University of California, Irvine, CA
Associate Dean's Fellowship **Spring 2013**

Pedagogical Fellowship **2012- 2013**
A year-long fellowship in advanced college pedagogy and academic job preparation that involved over 100 hours of training, experience designing UCI's TA training program, training new UCI TAs, giving teaching consultations for TAs, reviewing applications and giving interviews for Pedagogical Fellowship applicants, and a monetary award.

RESEARCH INTERESTS

Musical tonality and rhythm perception, visual and auditory multimodal binding and auditory localization.

TEACHING EXPERIENCE

University of California, Irvine, CA
Teaching Assistant – to Professor Charles Chubb in "Writing about Memory" **2010, 2011, 2012, 2013, 2014**
Met with students in office hours and upon request, graded essays, and collaborated on grading midterms and final exams. Led a lecture on writing style (2014).

Teaching Assistant – to Professor Barbara Sarnecka in "Developmental Psychology" **2012, 2014**
Developed curriculum for and led weekly discussion classes (2012), met with students in office hours and upon request, wrote exam questions (2012), and graded essays.

Teaching Assistant – to Professor Charles Chubb in "Principles of Learning Theory" **2011, 2013**
Led a lecture in collaboration with the other Teaching Assistant (2011), met with students in office hours and upon request, and collaborated on grading the midterm and final exams.

Teaching Assistant – to Professor Virginia Richards in "Perception and Sensory Processing" **2012**
Met with students in office hours, wrote homework and exam questions, and collaborated on grading midterms and final exams. Guest lecture on Music Perception.

Teaching Assistant – to Professor Christine Lofgren in "Introduction to Psychology" **2011**
Developed curriculum for and led weekly discussion classes, met with students in office hours and upon request, and collaborated on grading exams.

Teaching Assistant – to Professor Kourosh Saberi in "Experimental Psychology" **2010**
Developed curriculum for and led weekly labs, met with students in office hours and upon request, and graded written work and exams.

Teaching Assistant – to Professor Chuck O'Connell in "Intro to Sociology" **2010**
Developed curriculum for and led weekly discussion classes, met with students in office hours and upon request, and graded written work and exams.

Coastline Community College, Costa Mesa, CA
2012 Guest lecture on Memory for Instructor Meghan Goldman in "Intro to Psychology"

PEDAGOGICAL PRESENTATIONS

- Designing and leading the 1.5 day TA Professional Development Program for new Cognitive Science and Institute of Mathematical & Behavioral Science TAs (2012 & 2013)
- School of Social Sciences TA Training Workshops: Instructional Design: "Transforming Concepts into Effective Lesson Plans" (2012), "Grading Essays Successfully: Rubrics and Feedback" (2013), "Using Your Time Wisely: Time Management in Teaching and Work-Life Balance" (2013)
- Led weekly meetings for Undergraduate Research Mentorship group in the Chubb-Wright Lab (Fall 2012)

PUBLICATIONS & RESEARCH PRESENTATIONS

Chubb, C., Dickson, C., Fagan, C., Dean, T., Mann, D., Wright, C., Silva, A, Guan, H. Bimodal distribution of performance in discriminating major/minor modes. Journal of the Acoustical Society of America (2013)

Mann, D., Chubb, C. (2011, May). Binding Brightness and Loudness in Dynamic Audiovisual Displays. Poster presentation at the Vision Sciences Society 11th Annual Meeting, Naples, Florida.

Mann, D., Chubb, C. (2012, May). The temporal resolution of binding brightness and loudness in dynamic random sequences. Talk at the Vision Sciences Society 12th Annual Meeting, Naples Florida.

UNIVERSITY INVOLVEMENT

- Cognitive Neuroscience of Meditation Club at UC Irvine (2012-present)
- Real Food Challenge at UC Irvine member since 2009. Club Development Officer (2012-2013)- offered leadership training as well as organized social events for the club

COMPUTATIONAL SKILLS & LANGUAGES

- MATLAB Programming
- SPSS statistical software
- English – native language
- Spanish – speak, read, and write with basic competence.

PROFESSIONAL MEMBERSHIPS

- Vision Sciences Society (2011-2012)

ABSTRACT OF THE DISSERTATION

Processing Stimuli over Time: Musical Modes and Audiovisual Binding

By

Daniel Mann

Doctor of Philosophy in Psychology

University of California, Irvine, 2014

Professor Charles Chubb, Chair

This thesis covers three experiments related to processing rapid sequences of auditory and visual stimuli. Experiment 1 builds on the discovery that 70% (30%) of listeners perform near chance (perfectly) in classifying rapid sequences of tones (tone-scrambles) as major vs minor. Experiment 1 investigated the relationships between performance in various musical tasks, including the major/minor tone-scramble task. Skill in (1) judging the direction of pitch-change between two successive tones and (2) detecting the presence of an out-of-scale note in a melody were necessary but not sufficient for skill in classifying major vs minor tone-scrambles. These results suggest that skill in classifying major vs minor tone-scrambles requires a cognitive asset beyond those required for the interval-direction and scale-violation tasks. Experiment 2 tested how rhythm and pitch interact to control perceived majorness vs minorness. Participants classified three different types of tone-scrambles as major vs minor. All comprised 15 tones. In one condition, tone-scrambles had no rhythmic variation; in a second condition, every 5th tone was twice as long as the other 12 tones; in a third condition, every 5th tone was as long as a standard tone but was followed by a rest. Rhythmically accentuated tones influenced

judgments both more strongly and differently than unaccentuated tones. Moreover, the final tone influenced judgments differently than either standard tones or other rhythmically accentuated tones. Strikingly, when its final tone was a tonic, a tone-scramble was substantially more likely to be judged as "major." Experiment 3 explored how people can use top-down attention to bind information about brightness and loudness. Participants strove to classify rapid streams of disks varying in brightness presented simultaneously with noise-bursts varying in loudness in accordance with different attention instructions. Participants were able to attend to loudness only and ignore variations in brightness, but they had more trouble attending to brightness only and ignoring loudness. The various attention filters achieved by participants demonstrated that top-down attention can powerfully modulate the binding of loudness and brightness in dynamic displays.

CHAPTER 1: The ingredients for major/minor mode discrimination: pitch-height sensitivity, detecting a scale-violated melody, and more.

Abstract

Composers of Western music use major and minor modes to convey emotion assuming that nearly everyone is sensitive to the difference between modes. Chubb et. al 2013 discovered approximately 70% of people cannot discriminate major/minor modes in rapid sequences of tones or tone-scrambles. The remaining 30% achieved near-perfect performance in the major/minor discrimination task. Years of musical training only moderately correlated with performance in this task. The present study investigated what auditory abilities are necessary and/or sufficient to be sensitive to the major/minor discrimination task. Four tasks were tested: the Chubb et. al (2013) major/minor task, a pitch-height-comparison task, a scale-violated melody detection task, and a pitch-memory task. Additional data was collected regarding native language, musical experience and start of musical training. Native speakers of tonal languages performed better at the pitch-height comparison task. Musical training was moderately correlated to major/minor task performance, but start of musical training did not correlate with any of our tasks. The abilities to compare pitch-height and detect scale-violated melodies were necessary but not sufficient for major/minor task ability. These results provide part of the major/minor sensitivity puzzle, but there are certainly other features that separate the low- and high-sensitivity major/minor task groups.

The difference between major and minor musical modes is fundamental to Western musical theory and practice. Major and minor musical modes are purposefully used in music to establish particular moods; major modes are considered “happy” sounding and minor modes are considered “sad” sounding. These perceived emotional qualities of the musical modes are mysterious because there is no evident reason why certain combinations of notes should connotate emotions without a clear reference to anything with emotional content. Contrary to the expectations of music theorists, most people cannot discriminate major/minor modes even amongst musicians (Halpern 1984, Halpern 1998, Leaver & Halpern 2004).

Chubb et al. (2013) used a new class of stimuli called *tone scrambles* to show that major versus minor musical modes are not as clearly distinct for many listeners as music theory presupposes. The tone scrambles were composed of a rapid, randomly ordered sequence of pure tones drawn from either a major or minor scale. The major (or minor) tone scramble is composed of equal numbers of the low tonic, the major (or minor) third, the dominant and the high tonic. The main result of the Chubb et al. paper was a strikingly bimodal distribution of task performance in which the modes centered around chance and perfect performance. This reflected the existence of a distinct high-performing group and low-performing group in this major/minor discrimination task. Interestingly, Chubb et al. found only a modest correlation between years of musical training and performance in classifying major vs minor tone scrambles. Interestingly, many of the listeners in the low-performing group of the Chubb et al. study did express that music was important to them. Many participated in musical groups of various sorts and/or spent a great deal of time listening to music. It is unclear how and why these groups are separated.

It is possible that low-performers in the major/minor task are analogous to colorblind observers in that they lack a dimension of musical sensitivity that high-performers possess. Under this scenario, high-performers would have a generally richer experience of music than low-performers, and the low-performers might have a deficit in their auditory processing that would limit their musical pursuits. This is not the only possibility, however. Intervallic relationships and rhythmic structure establish much of the feel/emotional quality of music (e.g., Jackendoff & Lerdahl, 2006; Krumhansl, 2002; Lerdahl, 2009). By design, the tone scrambles used in the Chubb et al. study are devoid of all of the higher-order structural elements of actual music that enable listeners to extract its emotional meaning. It is possible that high-performing listeners differ from low-performing listeners in the major/minor task only in being able to extract musical meaning from such structurally impoverished musical stimuli. Under this scenario, participants who perform poorly at the tone-scramble classification task can experience the full emotional meaning of actual music without any issues.

The Present Experiment

This project investigates the relationship between major/minor mode sensitivity and various other sorts of pitch & musical abilities. In addition to the major/minor task from Chubb et al, listeners will be tested in three other tasks. One will test pitch memory, another will test the ability to compare pitch height, and the third will test the ability to detect a scale-violated note in a melody.

We will assess the dependencies between performance in the major/minor task with performance in the other three tasks. We say (1) task A is *necessary* for task B if any listener who performs poorly at task A also performs poorly at task B and (2) task A is *sufficient* for task B if any listener who performs well at task A also performs well at task B. The primary questions of interest are: (1) Are any of the other three tasks necessary and/or sufficient for the major/minor task? And (2) Is the major/minor task necessary and/or sufficient for any of the other tasks?

We shall also be alert for higher order dependencies. For example, it may turn out that any listener who performs poorly at both the pitch-memory task and also the scale-violated melody-comparison task will also perform poorly at the major/minor task. Such a result would suggest that skill in at least one of the pitch-memory task or the scale-violated melody-comparison task is necessary for a listener to demonstrate skill in the major/minor task. By clarifying the conditional dependencies that hold between these four tasks, we hope to gain insight into the functional architecture of major/minor mode sensitivity.

In addition, we will investigate the relationship between task performance and three other variables. We will ask participants to report (1) their years of musical training, (2) the age at which their musical training (if any) began, and (3) their native language. Years of musical training is an interesting factor to analyze since there were only moderate correlations with major/minor task performance in the Chubb 2013 paper. There is evidence that development of absolute pitch (AP) depends on receiving appropriate musical training during a critical acquisition period before about 9 years of age (Miyazaki & Ogawa, 2006; Russo et al. 2003; Sacks, 2007). This raises the possibility that skill in the pitch comparison task may show some dependency on the age at which musical training began.

Previous findings show that native speakers of tonal languages are more sensitive to pitch-height than native speakers of non-tonal languages (Bidelman, 2013; et al., 2006;

Giuliano et al., 2011). Due to these results, we expect that native speakers of tonal languages are likely to perform better in the pitch-height comparison task than speakers of non-tonal languages. A key question is: Do native speakers of tonal languages tend to perform better in the major/minor task than speakers of non-tonal languages? Such a finding would support the claim that skill in the major/minor task depends on exposure to the right sort of training early in life.

We expect a bimodal distribution of performance in the major/minor task (replicating the results of Chubb et.al (2013)), with more than half of our participants performing near chance and the rest performing near perfectly.

The major tone-scrambles differ from the minor ones in containing 8 B's instead of 8 B-flats. If a listener were able to (1) remember the difference between a B and a B-flat and (2) determine whether a given tone-scramble contained B's vs B-flats, then he/she would be able to perform the major/minor task successfully. If such a strategy predominated in the major/minor task, then we might expect skill in the pitch memory task to be necessary for skill in the major/minor task. However, this memory-based strategy does not seem to predominate among listeners skilled at the major/minor task. The memory-based strategy hinges on detecting the slight difference in pitch-height between the B's vs the B-flat's. However, skilled listeners do not experience the qualitative difference between major vs. minor tone-scrambles as a subtle difference in average pitch height; rather, the difference they experience (which leads them to say that major tone-scrambles sound "happy" whereas minor ones sound "sad") seems to be driven not by the B's vs the B-flats in isolation but rather by the intervals formed between the tonic G's, the dominant D's and the

B's vs the B-flat's. We therefore do not expect skill in the pitch memory task to be necessary for skill in the major/minor task.

Next, we anticipate that skill in the pitch-height-comparison task ability will be necessary for the pitch-memory task. If a listener is unable to discern whether a tone Y played immediately after a given tone X is higher or lower than X, it seems unlikely that the listener will be able to adjust the second tone Y to match the remembered tone X. Also, if the pitch memory strategy were employed in the major/minor task, skill in pitch-height-comparison would also be crucial to success in the major/minor task since the listener must be able to discern the difference between a B-flat and a B to use this strategy.

The scale-violated melody-comparison task (borrowed from Peretz et al., 2003) is not really a test of memory. Although the listener is asked to judge whether two successive melodies are identical or different, the first melody presented to the listener on a given trial always obeys the standard structural rules of western music, and whenever the second melody differs from the first, it differs in a single note that violates these rules by departing dramatically from the diatonic scale established by the melody. Thus, the scale-violated melody-comparison task is not a test of memory so much as a test of the sensitivity of a listener to departures from the musical context (i.e., the mode) that has been established by the melody.

With this said, it is important to recognize that the establishment of a musical context is itself a process that requires memory. A sense of (1) the scale used by the melody and (2) the notes emphasized as centering the melody can only be accrued as the melody unfolds in time. Does this process use the same mnemonic resources as the pitch-memory task? If so, then we might expect skill in the pitch-memory task to be necessary for skill in the scale-violated melody-comparison task.

We anticipate that a listener who cannot sense the mode of a melody will also have difficulty discriminating major from minor modes. Therefore, we expect listeners who perform poorly in the scale-violated melody-comparison task to also perform poorly in the major/minor task. On the other hand, there may exist listeners who (1) are able to sense the "clunker" notes that differentiate ill-formed melodies from their well-formed counterparts in the scale-violated melody-comparison task, yet who (2) cannot tell the difference between major vs minor ton-scrambles. After all, the notes occurring a tone-scramble—either a major one or a minor one—are all drawn from a single diatonic scale.

Method

Participants

112 undergraduate students were recruited from the Social Science Human Subjects Pool at the University of California, Irvine. Participation in the experiment was awarded with extra credit applied to one of their courses.

Apparatus

Participants worked on one of three PCs running Windows 7. The stimulus presentation and data collection were managed by a MATLAB program. All of the auditory stimuli were presented over headphones adjusted for comfort by each listener individually.

Conditions & Procedure

Each listener was tested in four auditory tasks: (1) the **Major/Minor** classification **Task** used by Chubb et al., 2013, (2) a **Pitch-Memory Task**, (3) a **Pitch-Height-Comparison Task**, and (4) a **Scale-Violated Melody-Comparison Task**.

Major/Minor Task: The stimuli in this task were created identically to the tone scrambles from Chubb et al (2013). The tone scrambles were each composed of 32 tones from the G_5 to G_6 octave. Each scramble contained 8 G_5s, 8 Ds, 8 Bs for the major scrambles (or B-flats for the minor scrambles), and 8 G_6s. In music theory terms, each stimulus contained 8 low tonics, 8 dominants, 8 major (or minor) thirds, and 8 high tonics. Each tone was 65ms long with a raised-cosine window to prevent the clicking sounds that would occur with abrupt onsets and offsets of each tone. The entire tone-scramble lasted approximately 2 seconds.

Participants first heard 8 sample stimuli. The examples alternated between major and minor tone-scrambles. When each example played, it was labeled as either "happy (major)" or "sad (minor)." After hearing the examples, the participants were asked to classify each tone scramble they heard as either "happy (major)" or "sad (minor)." The participant entered his/her response by pressing either a "1" or "2" on the keyboard. Trial-by-trial feedback ("correct" or "incorrect") was provided to enable participants to optimize their classification strategy. Participants ran 4 blocks of 50 trials each. A message after each block displayed the percent correct achieved by the participant in that block.

Pitch-Memory Task: Each stimulus was a 500 ms tone ranging from 300 – 2,000 Hz with the same raised-cosine window as the tones in the Major/Minor task. For each of 20 trials, the stimulus tone was played followed by a 2 second pause before participants could respond. After the pause, participants adjusted a slider to match to the stimulus tone in pitch. Each time the slider was adjusted, a new 500ms tone (again ranging from 300- 2,000 Hz) was played. Once the participant felt confident in their selection of the matching response tone, they clicked a button to advance to the next trial. Trial-by-trial feedback was displayed as the percent-of-an-octave difference between the stimulus and response tones. In other words, an entire octave difference was 100% off from the stimulus tone. An example feedback statement was “You were 5% above (or below) the original tone.” This line of feedback was displayed for 2 seconds followed by a 1 second pause before the next stimulus was played.

Pitch-Height-Comparison Task: In this task, participants were asked to judge whether the second of two tones was higher or lower in pitch than the first tone. Each tone had a duration 500ms long ranging from 300-2,000 Hz. There was a 2 second pause between the two tones. Before making any judgments, participants listened to 4 examples of tone-pairs. There were two examples in which the second tone was higher in pitch and two examples in which the second tone lower in pitch. After hearing the examples, participants began the actual experiment: on each trial, a pair of tones was played; then participants responded with a “1” if the first tone was higher in pitch than the second or a “2” if the second tone was higher in pitch than the first tone. Trial-by-trial feedback indicated whether or not the participant was correct or incorrect.

This task used two interleaved “3-down, 1-up” staircases to adaptively adjust the difficulty of the pitch comparison by decreasing (increasing) the difference between the two tones to make the task harder (easier). Each participant ran 2 blocks of 50 trials for this task.

Scale-Violated Melody-Comparison Task: This task was acquired with permission from the Montreal Battery of Evaluation of Amusia (MBEA) (Peretz et al., 2003). In Peretz et al (2003), this particular task was called the “scale-violated condition.” Performance in this task was highly correlated with all the melodic tasks in the MBEA. Participants were asked to judge whether two melodies, $melody_1$ and $melody_2$ were the same or different. All of the $melody_1$'s used in these stimuli were well-formed melodies typical of western music as opposed to random sequences of tones. If $melody_2$ differed from $melody_1$, then $melody_2$ contained one note that was out of the scale established by $melody_1$ and the remainder of the notes in $melody_2$. Each of $melody_1$ and $melody_2$ lasted approximately 5 seconds.

On every trial, the participant heard a warning beep, then $melody_1$, then a two second pause, then $melody_2$; the participant then responded with a "1" if the melodies sounded the same or a "2" if the melodies sounded different. Participants ran one block of 31 trials in this task. The block contained equal numbers of trials in which $melody_2$ was identical to vs different from $melody_1$ plus one catch trial in which the $melody_2$ consisted of random notes. If the participant missed this catch trial, then their data will not be considered (this procedure adheres to rule used in the original MBEA).

Procedure. Each participant completed all four tasks in the order assigned by a 4x4 Latin square for each computer. Participants were allowed to adjust the volume of the stimulus presentation to be comfortable. Participants were allowed to take breaks at any time, and they were asked to stay quiet with no humming (humming could have been used to cheat in the pitch-memory task). For each participant, the entire experiment took approximately 40 to 50 minutes. Prior to testing, each participant reported their native language, years of musical experience, and the age at which they began musical training (if any).

Results

Dependent Variables

Major/Minor & Scale-Violated Melody-Comparison Tasks. Instead of focusing on percent correct, we will calculate d' from Signal Detection Theory (Green & Swets, 1966) for the Major/Minor and Scale-Violated Melody-Comparison Tasks. The d' measure gives a purer estimate of stimulus discriminability than percent correct (which fails to take into account differences in the decision criteria used by different participants):

$$d' = \Phi^{-1}\left(\frac{Hits}{Hits + Misses}\right) - \Phi^{-1}\left(\frac{False\ Alarms}{False\ Alarms\ + Correct\ Rejections}\right) \tag{1.1}$$

where Φ^{-1} is the inverse of the cumulative Gaussian distribution.

Pitch-Memory Task. The data for the Pitch-Memory Task is a distribution of differences between the stimulus tone and the response tone in which participants selected as their match for each trial. As mentioned above, these differences will be calculated as percent of an octave above or below the stimulus tone. We will observe the RMSD for (Root Mean Squared Deviation) of the response from the target response. The RMSD metric is described as:

$$SD = \sqrt{\frac{1}{t}\sum_{k=1}^{t} d_k^2} \tag{1.2}$$

where t is the number of trials and d_k is the percent of an octave difference between the stimulus and response tone on trial *k*. The first two trials were considered training trials, so 18 trials were analyzed for each participant.

Pitch-Height-Comparison Task. The Pitch-Height-Comparison task analysis will require fitting psychometric functions for each participant. We will use a Weibull function fitting procedure to estimate the absolute value of the pitch difference for which the participant would achieve 82% correct in the pitch-height-comparison task. The cumulative distribution of the Weibull function is shown below:

$$\Psi = min + (max - min) * (1 - e^{-\left(\frac{x}{A}\right)^{B}}) \tag{1.3}$$

where min is the minimum probability correct, max is the maximum probability correct, x is the absolute value of the difference between the two pitches presented on a given trial, A is the threshold parameter, and B is the shape parameter. In the current case, we take min = 0.5 (the probability of a correct response by guessing) and max = 0.98 (to cover the possibility of a finger error even when the participant knows the correct response). The two free parameters of the Weibull function will be estimated using a Markov Chain Monte Carlo (MCMC) procedure in MATLAB. We will run 10000 samples of the MCMC procedure with 5000 trials of burn-in.

Major/Minor Task Replication

Performance in the Major/Minor Task replicated the bimodal result found in Chubb et. al (2013) (see **Figure 1.1**).

Figure 1.1 *Major/Minor Task Replication.* In the last 50 trials of the Major/Minor Task, performance is bimodal the proportion correct (left) figure. There is an artificial peak at d'=3.5 because d' estimates above about 3.5 cannot be accurately estimated by our data. Thus, we imposed is a ceiling for d' scores at 3.5.

Major/Minor Groups

The remainder of the results section will use the following abbreviations: **MM** = Major Minor Task, **PHC** = Pitch-Height-Comparison Task, **SVMC** = Scale-Violated Melody-Comparison Task, **PM** = Pitch-Memory Task.

For the following analyses of performance between tasks, we used the data from 103 participants. 9 participants were removed from this dataset because they either were flagged on the catch trial in the SVMC (4 participants) or there was extremely poor performance that suggested they were not paying attention in at least one of the tasks (5 participants).

One of our main questions was if people in the low- versus high-sensitivity group of Major/Minor Task (MM) would differ in their performance in the other 3 tasks. Participants with $d' < 2$ were placed in the low-sensitivity group, and participants with d'

>= 2 were placed in the high-sensitivity group. A d' score of 2 translated to approximately 84% correct. By this group assignment, 73 participants were in the low-sensitivity group and 30 participants were in the high-sensitivity group. **Figure 1.2** shows the difference between low- and high-sensitivity MM groups.

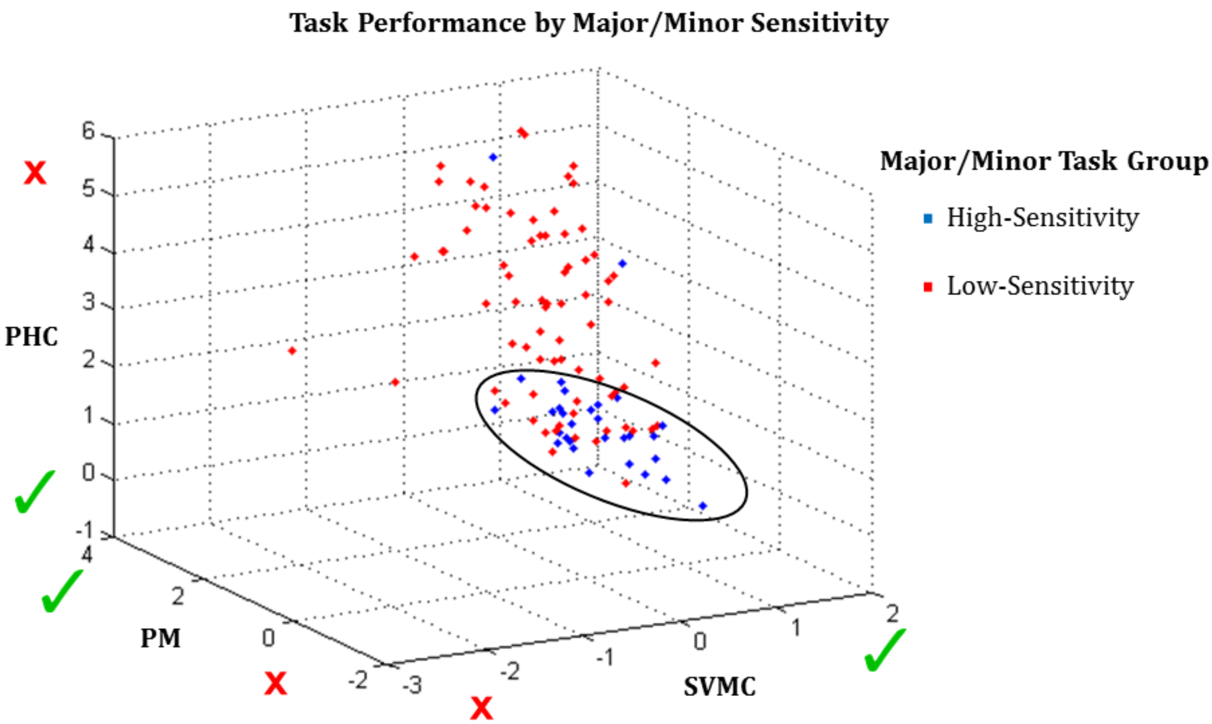


Figure 1.2 *Task Performance Grouped by Major/Minor Sensitivity.* This plot shows distinctly different 3-dimensional clouds for our data separated by low- and high- sensitivity to the Major/Minor Task (MM). For each axis, the red X indicates poor performance and the green check-mark indicates good performance. Each axis was natural-log transformed to make the distributions more normal. The Scale-Violated Melody-Comparison Task (SVMC) axis plots d' scores. The Pitch-Memory Task (PM) axis plots the RMSD of the distribution of the differences between stimulus and response tones. The Pitch-Height-Comparison (PHC) Task axis plots the threshold of accurate pitch-height-comparisons. The two high-sensitivity MM participants (blue dots) outside of the black circle are significant univariate and multivariate outliers.

In **Figure 1.2** we find that the high-sensitivity MM group performs well on the Scale-Violated Melody-Comparison Task (SVMC) and the Pitch-Height-Comparison Task (PHC), but this group has a wide variation of Pitch-Memory Task (PM) ability. The low-sensitivity MM group shows a wide range of ability for the other 3 tasks.

A one-way multivariate analysis of variance (MANOVA) was run to determine the effect of major/minor sensitivity on performance in the other three tasks. Two groups of major/minor sensitivity were assessed: high-sensitivity and low-sensitivity. The three dependent variables were performance on the PHC, PM, and SVMC.

Initially, our data failed some assumptions of the MANOVA. To begin with, the SVMC was normally distributed, as assessed by Shapiro-Wilk test ($p > .05$), but the data from the PHC and PM were not normal- Shapiro-Wilk test ($p<.001$). To make the PHC and PM datasets more normal, a natural-log transformation was performed. After this transformation, there were two outliers that were both univariate and multivariate, as assessed by boxplot and Mahalanobis distance ($p < .001$), respectively. These outliers were removed. There was no multicollinearity (PM/PHC $r = .447$, $p < .001$; SVMC/PHC $r = -.328$, $p = .001$; SVMC/PM $r = -.189$, $p = .059$). Our test violated the assumption of homogeneity of variance-covariance matrices, as assessed by Box's M test ($p < .001$). Even after normalizing the DVs with natural-log transformations and removing outliers, and the MANOVA revealed the same significant result as a MANOVA using original data set (untransformed, including outliers). Since the MANOVA test is robust to violations of many of these assumptions, we report the results from the MANOVA using the original data set.

The differences between the low- and high- MM sensitivity groups on the combined dependent variables was statistically significant, $F(3, 99) = 5.476$, $p = .002$; Pillai's Trace = .142; partial $\eta 2 = .142$. The Pillai's Trace test was used because it is more robust and recommended for unequal sample sizes and a statistically significant Box's M result.

Testing the violation of normality: An independent samples Mann-Whitney U test was performed on the untransformed data since it has weaker assumptions (e.g. no assumption of normality and outliers are okay) than t-tests; this tests the null hypothesis that the distributions are the same across the two groups. The Mann-Whitney U test gave the same basic results of the independent-samples t-tests listed below; thus, we report the results from the independent-samples t-tests.

Data are mean ± standard deviation unless otherwise stated. Follow-up unequal variance t-tests showed that **PHC thresholds were lower (better) for high-sensitivity MM participants** (7.50 ± 23.51) **than low-sensitivity MM participants** (39.57 ± 57.66), a statistically significant difference of 32.07 percent of an octave (95% CI, 16.20 to 47.94), $t(101) = 2.941$, $p < .001$. **SVMC d's were higher (better) for high-sensitivity MM participants** (2.64 ± .48) **than low-sensitivity MM participants** (2.09 ± .84), a statistically significant difference of -.559 (95% CI, -.87 to -.22), $t(90.258) = -3.339$, $p < .001$. These two-tests assumed unequal variance because they failed to reject the null hypothesis of Levene's test of equal variances ($p < .05$). The PM satisfied the equal variances assumption, and the t-test revealed that **PM scores were *not* significantly different across MM sensitivity groups**: $t(101) = 1.053$, $p = .295$. Hypothesis testing for these t-tests were evaluated with a Bonferroni adjusted $\alpha = .0167$.

Dependencies Between Tasks. Although our tests reveal that there are statistically significant differences between the means (and distributions) of our 3 tasks grouped by major/minor sensitivity, there is an interesting overlap of groups in the black circle in

Figure 1.2. This overlap suggests dependencies rather than distinct linear relationships between our tasks. By plotting performance in task A as a function of performance in task B, we can see whether A is necessary for B (i.e., whether any listener who performs poorly in A also performs poorly in B) or whether A is sufficient for B (i.e., whether any listener who performs well in A also performs well in B).

Our data suggest the following dependencies:

1. Pitch-Height-Comparison Task ability and Scale-Violated Melody-Comparison Task ability are both necessary but not sufficient for Major/Minor Task sensitivity.
2. Pitch-Memory Task ability is neither necessary nor sufficient for Major/Minor Task sensitivity.
3. Pitch-Height-Comparison is necessary but not sufficient for Pitch-Memory Task ability.

See **Figure 1.3** for scatterplots illustrating dependency 1 and see **Figure 1.4** for a scatterplot showing dependencies 3 & 4.

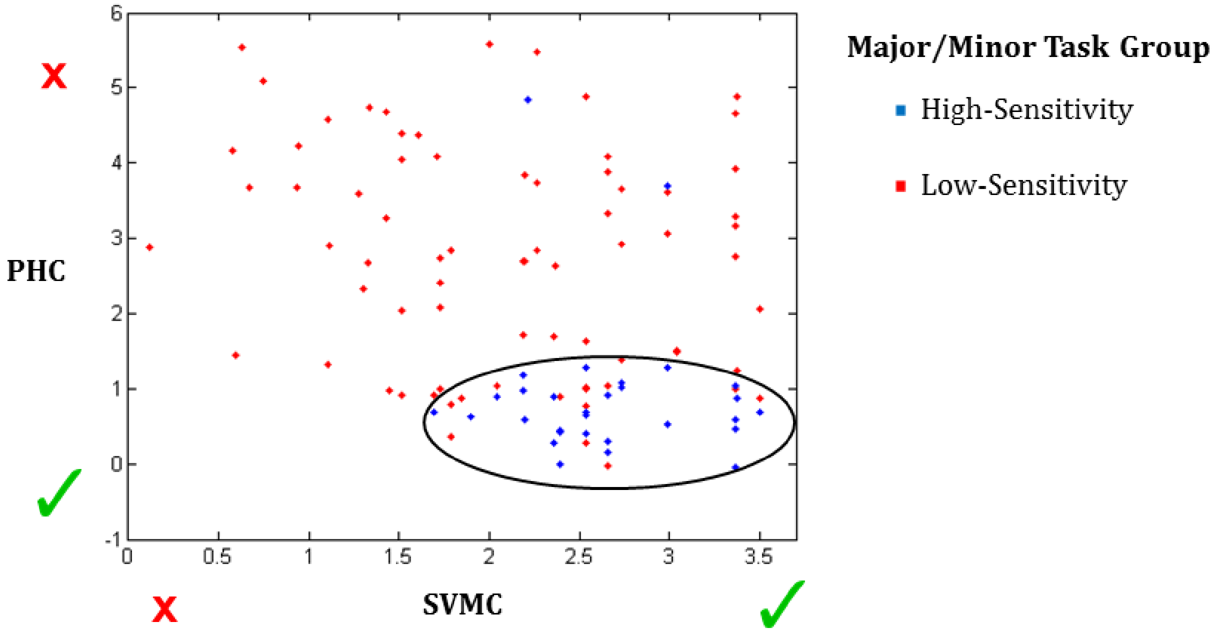


Figure 1.3 *SVMC and PHC Performance Separated by MM Group.* Results from the Scale-Violated Melody-Comparison Task (SVMC) and the Pitch-Height-Comparison Task (PHC) and are plotted and separated by the low- and high-sensitivity Major/Minor Task (MM) groups. For each axis, the red X indicates poor performance and the green check-mark indicates good performance. The PHC axis plots the threshold of accurate pitch-height-comparisons (natural-log transformed). The SVMC axis plots d' scores. The two high-sensitivity MM participants (blue dots) outside of the black circle are significant outliers for the PHC. This plot demonstrates that good performance on the PHC and SVMC are necessary to be in the high-sensitivity MM group, but they are not sufficient because there are high performing participants in both tasks that are not sensitive to the MM (the red dots in the black circle).

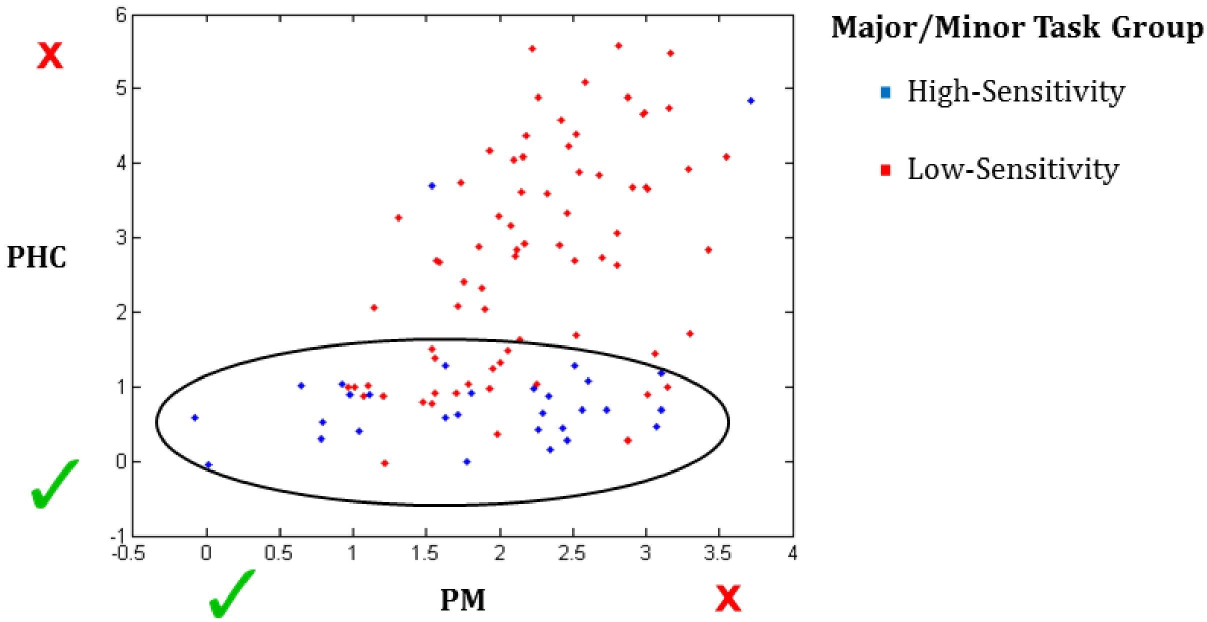


Figure 1.4 *PM and PHC Performance Separated by MM Group.* Results from the Pitch-Memory Task (PM) and Pitch-Height-Sensitivity Task (PHC) are plotted and separated by the low- and high-sensitivity Major/Minor groups. For each axis, the red X indicates poor performance and the green check-mark indicates good performance. Each axis was natural-log transformed to make the distributions more normal. The PM Task axis plots the standard deviation of the distribution of the differences between stimulus and response tones. The PHC axis plots the threshold of accurate pitch-height-comparisons. The two high-sensitivity MM participants (blue dots) outside of the black circle are significant outliers for the PHC. This plot demonstrates that good performance on the PM is neither necessary nor sufficient to be in the high-sensitivity MM group. Also, PHC ability is necessary but not sufficient for PM ability. There was a moderate correlation between PHC and PM for both log-transformed ($r = .456$) and original ($r = .357$) data.

Summary. The high-sensitivity Major/Minor (MM) group achieved significantly better Pitch-Height-Comparison (PHC) and Scale-Violated Melody-Comparison (SVMC) scores versus the low-sensitivity MM group. The Pitch-Memory Task (PM) performance was not different between MM groups. The scatterplot results suggest the following dependencies: PHC and SVMC are necessary but not sufficient for MM; PM is neither necessary nor sufficient for MM; PHC is necessary but not sufficient for PM.

Native Speakers of Tonal Languages

Our sample included 23 native tonal language speakers and 80 native speakers of non-tonal languages. The reported tonal languages were Chinese, Cantonese, Mandarin, Vietnamese, Thai, and Vietnamese.

Assumption checking: The Tonal Language and Non-Tonal Language groups for the PHC, MM, and PM datasets failed the normality assumption as assessed by Shapiro-Wilk test ($p < .05$). Independent-samples Mann-Whitney U tests gave the same general results as independent-samples t-tests, and since t-tests are generally robust to violations of the normality and outlier assumptions, we report the t-test results below.

Independent-samples t-tests were run to determine if there were differences in each of the MM, PHC, SVMC, and PM scores between native tone and non-tonal language speakers. Data are mean ± standard deviation unless otherwise stated. An unequal variance t-test showed that **PHC thresholds were lower (better) for native speakers of a tonal language** (8.88 ± 14.04) **than speakers of non-tonal languages** (36.36 ± 57.31), a statistically significant difference of 27.49 percent of an octave (95% CI, 13.51 to 41.46), $t(99.81) = 3.902$, $p < .001$. The following three t-tests had homogeneity of variance, assessed by Levene's test of equal variances ($p>.05$). **There was no significant difference between SVMC d's for native speakers of a tonal language** (2.31 ± .77) **and speakers of non-tonal languages** (2.23 ± .81), $t(101) = -.411$, $p = .682$. **There was no significant difference between MM d's for native speakers of a tonal language** (1.31 ± 1.22) **and speakers of non-tonal languages** (1.14 ± 1.41), $t(101) = -.503$, $p = .616$. **There was no significant difference between PM scores for native speakers of a tonal language** (10.13 ± 7.74) **and speakers of non-tonal languages** (10.97 ± 7.75), $t(101) = .459$, p =

.647. Hypothesis testing for these t-tests were evaluated with a Bonferroni adjusted α = .0125.

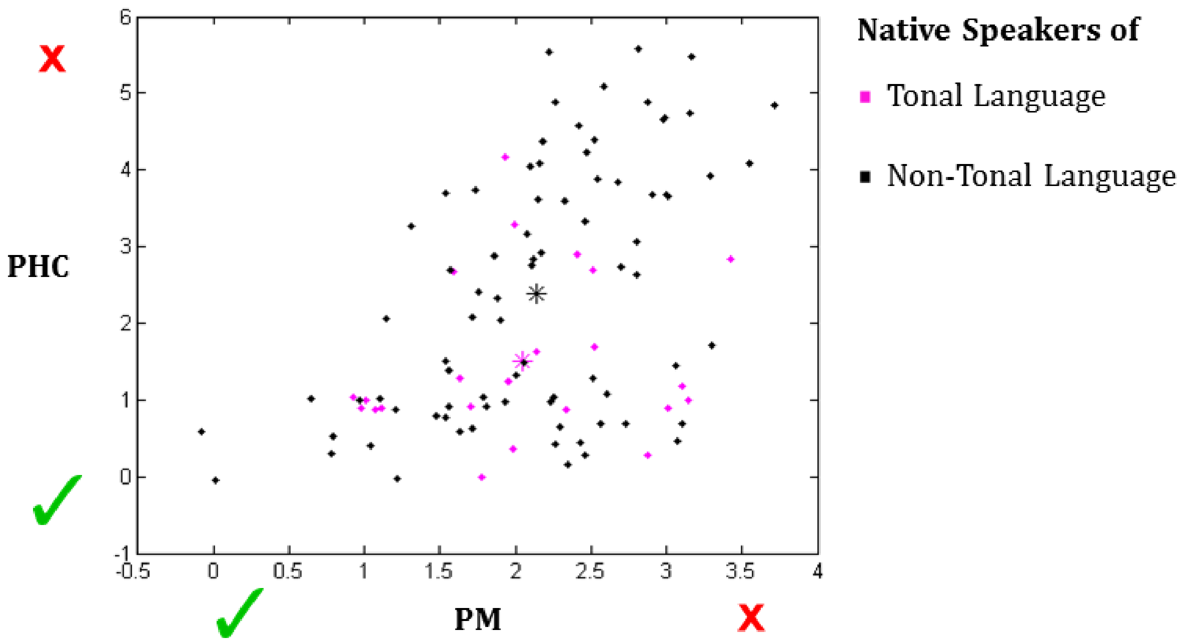


Figure 1.5 *PM and PHC Performance Separated by Tonal Language.* Results from the Pitch-Memory Task (PM) and Pitch-Height-Sensitivity Task (PHC) are plotted and separated by the Tonal Language and Non-Tonal Language groups. For each axis, the red X indicates poor performance and the green check-mark indicates good performance. Each axis was natural-log-transformed to make the distributions more normal. The PM Task axis plots the standard deviation of the distribution of the differences between stimulus and response tones. The PHC axis plots the threshold of accurate pitch-height-comparisons. The magenta asterisk is the mean score for native speakers of a tonal language and the black asterisk is the mean score for native speakers of a non-tonal language. There is a significant difference in the mean for the PHC dimension but not for the PM dimension.

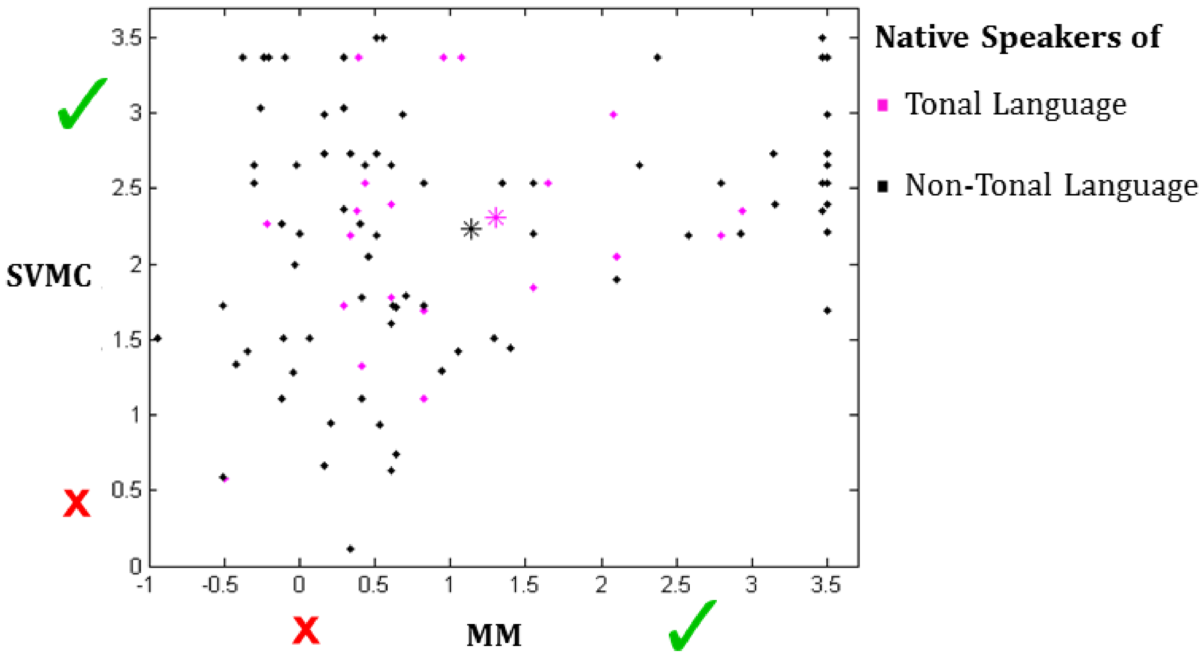


Figure 1.6 *MM and SVMC Performance Separated by Tonal Language.* Results from the Major/Minor Task (MM) and Scale-Violated Melody-Comparison Task (SVMC) are plotted and separated by the Tonal Language and Non-Tonal Language groups. For each axis, the red X indicates poor performance and the green check-mark indicates good performance. Each axis plots d' scores. The magenta asterisk is the mean score for native speakers of a tonal language and the black asterisk is the mean score for native speakers of a non-tonal language. There is no significant difference between these means on either dimension.

Summary. Native speakers of tonal languages on average had lower pitch-height discrimination thresholds than native speakers of non-tonal languages. There were no statistically significant differences between MM, PM, or SVMC scores for native speakers of tonal languages and native speakers of non-tonal languages.

Musical Training

Years of Musical Training. There was a moderate positive correlation between years of musical training and the MM, r = .469. There were small correlations between years of musical training and: PHC: *r* = -.230[1]; SVMC: *r* = .196. There was no significant correlation between years of musical training and PM[2].

Start of Musical Training. There were no significant correlations between the start of musical training and any of our four tasks.

Summary. Musical experience did not strongly correlate with performance in any of the 4 tasks. Musical experience moderately correlated with the Major/Minor Task sensitivity (replication of Chubb et. al 2013 result). Start of musical training did not correlate with any of our tasks. These results rejected our hypothesis that musical experience would correlate strongly with the Scale-Violated Melody-Comparison Task ability.

Predicting Major/Minor Group

A logistic regression was performed to ascertain the effects of task performance (PM, PHC, and SVMC), tonal language, and years of musical training on the likelihood that participants are in the high-sensitivity Major/Minor Group[3]. The PHC and PM scores were log transformed to remove outliers. The logistic regression model was statistically significant, $\chi^2(5) = 47.279$, $p < .0005$. The model explained 52.5% (Nagelkerke R2) of the variance in major/minor group and correctly classified 82.5% of cases. Of the six predictor variables only three were statistically significant: PHC, SVMC and years of musical training (as shown in **Table 1.1**). These results provide converging evidence for the differences found in the previous statistical tests.

---

[1] The correlation between years of musical training and the natural-log transformed PHC was moderate: r = -.381.
[2] Years of musical training and PM remained uncorrelated after a natural-log transformation of PM.
[3] Start of musical training was excluded from the logistic regression because there were many missing cases of people without any musical training.

	Wald	df	p-value
PVMC	3.947	1	.047*
Music Training	4.409	1	.036*
Log PHC	12.583	1	< . 001*
Log PM	.549	1	.459
Tonal_Lang(1)	.916	1	.338
Constant	3.438	1	.064

Table 1.1 *Predictor Variables in Major/Minor Group Assignment*. A logistic regression revealed that performance in the Pitch-Violated Melody-Comparison Task (PVMC), the Pitch-Height-Comparison Task (PHC), and years of musical training were statistically significant in predicting Major/Minor sensitivity group assignment.

Discussion

The present study replicated the Chubb et. al 2013 result in which the Major/Minor Task (MM) performance was bimodal with modes near chance (low-sensitivity) and perfect (high-sensitivity) performance. The high-sensitivity MM group had lower pitch-height-comparison thresholds than low-sensitivity MM group. The high-sensitivity MM group was also better at the Scale-Violated Melody-Comparison Task than the low-sensitivity MM group. There were no significant differences in the Pitch-Memory Task for high- and low-sensitivity MM groups. It is possible that the process of matching the pitch added enough difficulty to reduce the assessment of pitch-memory.

Pitch-height sensitivity and scale-violated melody-comparison ability were necessary to be in the high-sensitivity MM group. These two qualifications make sense because first, in order to discriminate the major/minor tone-scrambles participants must be able to discriminate major and minor 3rds (one semi-tone difference). Secondly, the Scale-Violated Melody-Comparison Task tested the ability to compare auditory sequences extended over a series of tones which is also crucial to make the major/minor judgments.

Interestingly, those two abilities did not guarantee (were not sufficient for) Major/Minor Task sensitivity. There were participants that performed well on both the Pitch-Height-Comparison Task and the Scale-Violated Melody-Comparison Task who were not sensitive to the Major/Minor Task. Thus, another factor (that we did not test for) must contribute to Major/Minor sensitivity.

Native speakers of tonal languages achieved significantly lower pitch-height-comparison thresholds compared to native speakers of non-tonal languages. Native speakers of tonal languages were able to accurately discriminate pitch-heights that were closer together compared to native speakers of non-tonal languages. This finding is in accordance with the existing literature on the relationship between speaking a tonal language and enhanced pitch discrimination (Bidelman, 2013; et al., 2006; Giuliano et al., 2011).

Music experience had a moderate relationship with Major/Minor Task performance as described by Chubb et. al 2013. There were small correlations with the pitch-height comparison thresholds and scale-violated melody-comparison task performance. Our logistic regression analysis demonstrated that years of musical training were a significant predictor for MM group assignment in our model. Start of training showed no significant correlations with performance in any of our tasks.

Pitch-Height-Comparison Task ability was necessary but not sufficient for the Pitch-Memory task. This result confirmed our simple hypothesis that in order to remember a pitch accurately, a person would need to be able to discriminate pitches well. The necessary but not sufficient relationship is understandable because Pitch-Memory Task did certainly require other skills than Pitch-Height-Comparison ability.

The Scale-Violated Melody-Comparison Task tapped into more than one cognitive ability. It was possible to do this task by either comparing the two sequences note-by-note in memory or listening for the out-of-scale note in the second melody. Of course participants may have been using some combination of these abilities to perform the task. This ambiguity is not of great concern because we can still assess whether major/minor discrimination ability is related to higher level musical abilities that require making judgments based on the relationships of groups of notes.

By better understanding the relationship between performance in the major/minor task and other auditory tasks, we may be able to discover a training regimen that can be used to heighten sensitivity to major/minor modes. In the Chubb 2013 paper, for the intermediate participants that scored an average of 60-90% across all 4 blocks, there was evidence of statistically significant improvement of performance between blocks 1 and 4. This fact demonstrates that, at least for some listeners, skill in the major/minor task is not fixed but rather may be acquired through training. Chubb et al (2013) also observed a large number of listeners who performed near chance in the major/minor task across all four blocks, suggesting that there may exist listeners for whom this task may be unlearnable. Pitch-height- and scale-violated melody-comparison training may assist major/minor discrimination, but there is definitely some other cognitive component necessary to make the major/minor discrimination that was not revealed by this experiment. Additional cognitive tasks must be tested to clarify the nature of the cognitive asset that is critical for classifying major vs minor tone-scrambles.

CHAPTER 2: End on a high note: Resolution changes the influence of pitch when discriminating major and minor modes

Abstract

Rhythm and pitch interact in their influence on the perceived emotional quality of short melodies (Schellenberg, et al., 2000). Major and minor tonalities are often associated with sounding "happy" and "sad" respectively. To investigate the role of rhythm in the major/minor mode discrimination, 3 participants classified major/minor sequences of tones with rhythmic manipulations. The sequences of tones or "tone scrambles" were composed of a random mixture of brief tones from a G major or minor mode (non-diatonic tones were also randomly included). Tone scrambles were rhythmically varied in three ways: the addition of (1) extended-notes, (2) rests, or (3) without rhythmic accents. Trial by trial feedback was provided to encourage participants to optimize their responses. A probit model was used to measure the impact of particular tone-types and their temporal positions on the major/minor discrimination. Rhythmic accents created by extended-notes created larger boosts of sensitivity to tones than the accents created by rests. The final tone created an accent that was more powerful than the local rhythmic accents (extended-notes and tones before rests) within a sequence of tones. The stability of the final note was related to its "major"/"minor" impact. Ending on the tonic made tone scrambles sound more "major," and for some participants, ending on unstable notes made tone scrambles sound more "minor." Rhythm did not simply amplify sensitivity; the unique impact of tones at the end of the sequence supported an inseparable model of rhythm and pitch interactions when discriminating major/minor sequences.

People produce periodic rhythms in many basic behaviors– infant sucking, rocking, walking, swimming, etc. These behaviors have periods between approximately 500ms to 1s (Fraise, 1982). Besides period behaviors, internal rhythm has been represented by spontaneous tempo, measured by the rate of moving a body-part (tapping a finger or palm, and swinging an arm or leg), and preferred tempo, measured by adjusting the rate of presentation of sounds. The average spontaneous tempo and preferred tempo is around 600ms (Fraise, 1982). The similarity in periodicity across these behaviors and preferences suggest an internal pulse, but people are also sensitive to external pulses since we naturally synchronize movements to rhythmic sounds. This synchronization happens spontaneously in infants as young as 1 year old (Fraise, 1982). This sensitivity and physical connection to periodic patterns is the cognitive foundation for the use of rhythm in music.

Musical rhythm is created by three types of accents- phenomenal, structural, and metrical (Lerdahl & Jackenoff, 1983). Of these three, phenomenal accents are the most physically concrete. A phenomenal accent is any physical feature of an instant of music that grabs the listener's attention- features such as jumps in pitch, chord changes, extended notes, and dynamics (Lerdahl & Jackenoff, 1983). Structural accents are defined by the way particular tones & intervals move from one another in a given musical key (Dawe et. al 1993, Lerdahl & Jackenoff 1983). Metrical accents are perceived through mental schemes that define "strong" and "weak" beats in a musical sequence (Dawe et. al 1993, Lerdahl & Jackenoff 1983). For example, in a 4/4 meter[4], beats 1 & 3 are typically considered the "strong" beats and beats 2 & 4 are the "weak" beats. A basic form of metrical accenting is observed when people imitate equal sounds of a clock as "tick-tock-tick-tock…" with extra

[4] Music with groupings, or bars, that consist of 4 quarter-note beats.

emphasis on the "ticks". In this case, the "ticks" are the strong beats, and emphasis is often created by increasing the volume in comparison to the "tocks." This volume increase demonstrates a metrical accent combined with a phenomenal accent, but this co-occurrence is not necessary to perceive metrical accents (Dawe et. al, 1993). The present experiment focused on manipulations of phenomenal and metrical accents.

Rhythm & Pitch

Rhythm and pitch typically interact in various music perception tasks. For example, Boltz (1989) found an interaction between rhythm and pitch for "completeness" ratings. The stimuli were folk tunes in which the final two tones were manipulated to be different scale degrees. Some of these combinations (e.g. the leading tone to tonic) were rated as particularly complete and other combinations (e.g. the tonic to leading tone) were rated as incomplete. The rhythms of the tunes were also manipulated to end on time, too early, or too late according to the established metrical structure. When the tunes ended uncomfortably *both* melodically and rhythmically (too early or too late), the incomplete ratings were unexpected. When both joint expectancies for melody and rhythm were violated, the incomplete ratings were more extreme than the combined impact of the single violations of melody and rhythm.

Many other experimental paradigms have demonstrated how variations in pitch affect rhythm perception and vice versa. In tasks where people compared two pitches within a sequence of tones, pitch was most accurately compared in rhythmically regular sequences versus rhythmically irregular sequences (Jones, Moynihan, MacKenzie, & Puente 2002, Jones, Johnston & Puente, 2006). In other words, the timing of tones influenced pitch judgments. Various experiments have shown that people struggled with recognizing a melody based on pitch when rhythmic variations were introduced (Jones & Ralston, 1991; Jones, Summerell, & Marshburn, 1987; Kidd, Boltz, & Jones, 1984). Variations in pitch can also affect the perception of tone duration (Boltz, 1992 as cited in Dawe 1995).

Motivations for the present study. In 1982, Jones, Boltz & Kidd found that pitch changes, in relation to melodic structure, were easier to detect at metrically accented locations in a sequence of tones that followed a predictable temporal pattern. This finding in part led to their Dynamic Attending Theory for complex auditory sequences (Jones & Boltz, 1989, see Drake 2000 for a review). The theory describes a model in which attentional energy is periodic based on many relative internal rhythms, or oscillators. Jones and Boltz claim that these oscillators are distinct from a "biological clock" because attention operates on relative timing and not absolute time.

This result raises the question of how periodic patterns can influence major versus minor mode discrimination. Since Jones & Boltz claimed that predictable temporal patterns can heighten sensitivity to melodic structure, the present study will investigate how musical rhythm influences tonal sensitivity in major/minor classifications.

Schellenberg, Krysciak, and Campbell (2000) had participants rate the emotional quality (happy, sad, scary) of melodies in which pitch and rhythm were manipulated. Rhythmic manipulations only had a significant effect in emotional quality ratings when pitch was also manipulated. The details of this interaction varied depending on each melody or musical context. Since rhythm can affect the emotional experience of short melodies, rhythm may influence the perception of the common "happy"/"sad" association with major/minor modes.

The Present Study

The previously discovered interactions of rhythm and pitch do not necessarily distinguish whether pitch and rhythm are *separable* or *inseparable* in various tasks. The present study will investigate the separability of rhythm and pitch in its influence on major/minor mode discrimination. The first possible model is the "separable model." This model describes that all notes basically exert the same pattern of major/minor influence with or without rhythm, but the magnitude of influence is amplified or attenuated by the rhythmic class of the note (how salient the note is made by the rhythmic structure). Next, there is the "inseparable model" of rhythm and pitch for major/minor classification. This model describes that notes exert different patterns of influence depending on the rhythmic class of the note.

The present study assessed major/minor mode discrimination using a variant of the major/minor task, developed by Chubb et. al (2013) (see Chapter 1 for details). The main research questions were the following: For people sensitive to the difference between major vs minor tone scrambles…

- does rhythm simply amplify or attenuate the influence exerted by particular tones in their major/minor impact (separable model)? Or is the pattern of influence intrinsically different for tones at particular rhythmic locations (inseparable model)?
- do phenomenal and metrical accents have an equal influence on participant's major/minor discrimination?
- which tones have the most influence in the major/minor discrimination?

The present study manipulated rhythmic structures in the tone-scrambles by adding extended-notes (phenomenal accents), rests (metrical accents), and comparing the results to a non-rhythmic control condition. We hypothesized that rhythm would interact inseparably with pitch in the major/minor classifications. We expected that extended notes, phenomenal accents, would be more salient than the metrical accents created by rests because of the extra signal that an extended note provides. We expected that tones will be weighted according to the major/minor notes defined by music theory because participants would receive trial-by-trial feedback to encourage this particular tone weighting.

This is the first study of rhythm and pitch to use linear process models estimated by Bayesian Markov Chain Monte Carlo sampling methods. Model comparisons will be evaluated with likelihood-ratio tests to determine the separability of rhythm and pitch in major/minor mode classification.

Methods

Participants

Three individuals participated in this experiment (3 males). Two of the participants were the authors, and the other was an undergraduate research assistant. All had self-reported normal hearing. Each participant gave informed written consent approved by the Institutional Review Board at the University of California, Irvine. All participants were highly sensitive (i.e. could achieve perfect performance) to the original major/minor task by Chubb et. al (2013).

Apparatus

The experiment was run on MATLAB on various computers. Participants sat in quiet environments in front of a computer while they listened to the stimuli over headphones. Each participant set the volume to a comfortable level for themselves.

Stimuli

The stimuli, or tone scrambles, were composed of pure tones from an equally tempered scale between G5 and G6- the low and high tonics of the scale. Because we included both tonics, there were 13 possible tones.

Each tone in a tone scramble was called a *pip.* 15 pips of 100ms duration each were presented in each tone scramble[5]. Each pip was presented at 50,000 samples per second. Each pip had a raised cosine window in which the onsets and offsets were ramped for 22.5ms. This created smooth transitions between each pip without clicks and pops.

The 15 pips in each tone scramble were defined by *note-count vector*. The note-count vector indicated the number of times each of the thirteen possible tones occurred in a particular tone scramble. For example, the tone scramble from note-count vector V = (2, 1, 0, 2, 0, 1, 1, 1, 0, 0, 2, 1, 4) would have 2 G_5s, 1 G♯, 0 As, 2 A♯s, 0 Bs, 1 C, 1 C♯, 1 D, 0 D♯s, 0 Es, 2 Fs, 1 F♯, and 4 G_{6s}. The 15 tones in a note-count vector were presented in random order for each tone scramble.

The difficulty was increased by modulating the note-count vectors to:

1. include more distractor tones, tones irrelevant to the major/minor discrimination

[5] The original tone-scramble experiment by Chubb et. al (2013) used stimuli that each had 32 tones, each with a duration of 65ms. The present study used 15 tones per stimulus, each with a duration of 100ms. The lengthened tones of the present stimuli created a more musical melody than the original, rapid tone scrambles.

2. have a less-clearly defined major/minor tonality. This was created by having nearly equal amounts of major and minor tones, so the entire scramble was just slightly major or minor.

Each level of difficulty had 24 possible note-count vectors (except for Difficulty 2 with 12 possible note-count vectors). There were 4 levels of difficulty. Difficulty level 1 had note-count vectors that only included the low and high tonics, the 3rds, 5ths and 6ths. In other words, for Difficulty level 1, there were no distractor tones (the tonic and 5ths help establish the scale). The difficulty levels 2-4 included varying amounts of distractor tones depending on the particular note-count vector. Since each tone-scramble was a random ordering of one of these note-count vectors, there was a high amount of variation between different tone-scrambles (even if they happened to have identical note count vectors). All the stimulus note-count vectors for each difficulty can be found in **Appendix A**.

Conditions

Three experimental conditions were distinguished by the rhythmic content of the stimuli. First, the *Extended-Notes* Condition used tone scrambles in which the 5th, 10th and 15th pips were twice as long as the other pips; the extended-notes were phenomenal accents. This created a 6/8 meter with 4 eighth notes (100ms each) and one quarter note (200ms) for each of 3 measures. Next, the *Rests* Condition used tone scrambles with 5 eighth notes (100ms) and one eighth note rest (100ms) for each of 3 measures. The accents were still on the 5th, 10th, and 15th notes in the rests condition, but these accents were metrical since they indicated phrase endings in the inferred meter. Lastly, the *Non-Rhythmic* Condition played 15 (100ms) pips. See **Figure 2.1** for rhythmic notations of each condition.

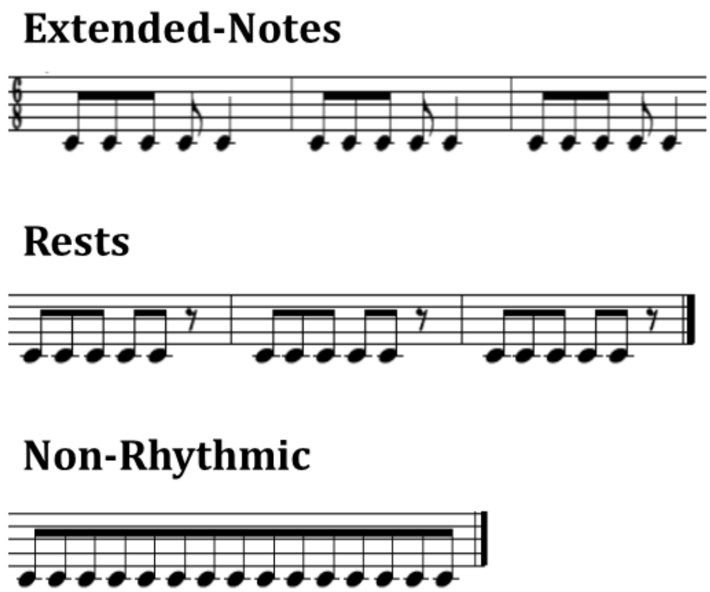


Figure 2.1 *Rhythmic Notation for Conditions.* These are the rhythmic notations for the Extended-Notes, Rests, and Non-Rhythmic Conditions. To focus on the rhythmic properties, this figure does not include sample note values of a tone-scramble.

In all conditions and difficulty levels, the feedback was derived from the *target function*. The *target function* defined whether the sequence was “major” or “minor” based on the number of major and minor 3^{rds} and 6^{ths} in the scale. The target function designated each major 3^{rd} and 6^{th} as +1 and each minor 3^{rd} and 6^{th} as -1, and all the other 9 possible tones were valued at 0. The sum of all these target function weights for each pip in a tone scramble defined whether the scramble was major or minor for purposes of giving the participant trial-by-trial feedback. See **Figure 2.2** for an example of how the target function was applied. If participants could weight each note exactly as the target function did, then they would get 100% correct on all trials. The trial-by-trial feedback encouraged participants to optimize their strategy based on the target function.

	G_5	G♯	A	A♯	B	C	C♯	D	D♯	E	F	F♯	G_6
Target Function (*TF*):	0	0	0	-1	1	0	0	0	-1	1	0	0	0
Note-Count Vector:	3	0	0	0	4	0	0	2	0	4	0	0	2

***Stimulus*:** B E B B G_6 B E G_5 E G_5 D G_5 G_6 D E

$\sum$ ***TF*(*Stimulus*)** = 1 + 1 + 1 + 1 + 0 + 1 + 1 + 0 + 1 + 0 + 0 + 0 + 0 + 0 + 1 = 8

Figure 2.2 *Target Function Example.* The target function, *TF*, is shown above with a sample note-count vector. The note-count vector was randomly ordered to form the tone-scramble, *Stimulus.* Then the target function, *TF*, was applied to the *Stimulus* and summed. Since this sum was positive, the tone-scramble was major. If the participant responded "Major" then they would get feedback on the screen that said "Correct." This target function was used for all conditions and levels of difficulty.

Procedure.

This experiment was a within-subjects design with 3 conditions. In each condition, participants completed 60 separate blocks of 50 trials for a total of 3000 trials each. The conditions were completed in the same order: Extended-Notes, Rests, then Non-Rhythmic.

Each trial began with hitting the "enter" key; then the tone scramble was presented over headphones. Then, the participants responded with either a "1" or "2" to indicate "Minor" or "Major" respectively, and feedback was always displayed on the screen ("Correct" or "Incorrect") after each trial.

In the Non-Rhythmic condition, participants responded from "1" to "4" where "1" meant "Very Minor" and "4" meant "Very Major." Feedback in the non-rhythmic condition was still based on the same major/minor rule. For example, if the stimulus was minor, then either a “1” or “2” response would be considered correct, and both the “3” and “4” responses would be incorrect.

Upon completing all 50 trials of each block, percent correct was displayed for that particular block. If percent correct was at 90% or higher, participants were encouraged to try a higher level of difficulty on the following block. Our participants most commonly used difficulty levels 2 and 3.

Modeling

Basic Model

We modeled the major/minor discrimination process with a probit model[6]. Before getting into the full model, we review the basic model logic.

In any given trial, we assumed participants would respond “Major” (versus “Minor” otherwise) if

$$\sum_{k=1}^{15} f(tone_k) * t(k) + Noise > Criterion \tag{2.1}$$

where f is the tone-weighting function (constrained to sum to 0 and to have sum of squared values equal to 1), t is the temporal weighting function (constrained to be nonnegative), and k is the position of each tone in the stimulus. f weights the influence of each tone on the participant’s major/minor decision. t assigns weights to the relative impact of each particular temporal position of each tone. The *Noise* is a standard normal random variable,

[6] Probit model is a general linear model with cumulative normal linking function.

and *Criterion* is a parameter that models the criterion selected by the participant to optimize performance.

Full Model

Extended-Notes and Rests Conditions. The basic model was separable in tone and time, but we investigated interactions by expanding the model. The full model included 3 tone-weighting functions- f_{final}, $f_{accented}$, f_{other}. f_{final} revealed the influence of each tone at the final pip in a stimulus (15), $f_{accented}$ gave the influence of each tone at the rhythmically accented pips within the stimulus (5, 10) and f_{other} was the tone-weighting function for all the other pip positions. These 3 separate functions allowed us to evaluate the influence of tones at each of these special locations.

With these added tone-weighting functions, in any given trial, we assumed participants would respond "Major" (versus "Minor" otherwise) if

$$\sum_{k=1}^{15} f_k(tone_k) \, * t(k) + Noise > Criterion \tag{2.2}$$

where

$$f_k(tone_k) = \begin{cases} f_{final}(tone_k), & k = 15 \\ f_{accented}(tone_k), & k = \{5, 10\} \\ f_{other}(tone_k), & otherwise \end{cases} \tag{2.3}$$

where k is the position of each tone in the stimulus. A graphic depiction of the model is shown in **Figure 2.3.** The constraints of the full model are listed in **Table 2.1**.

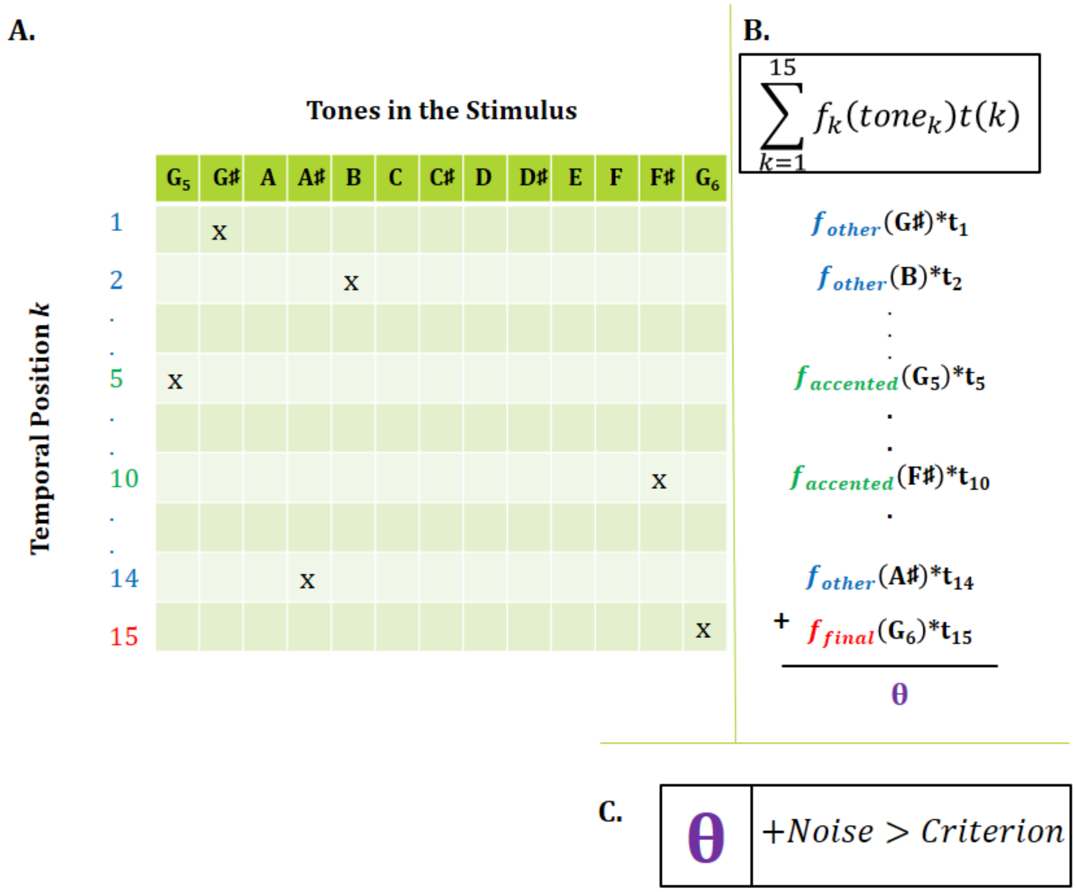


Figure 2.3 *Full Model Example*. **A.** This is a sample tone-scramble with various notes for each of the 15 temporal positions. This diagram has gaps in the rows just to save space. **B.** The appropriate tone-weighting function is applied according to the temporal position: f_{final} gives the tone-weight for the final pip; $f_{accented}$ gives the tone-weights at the 5th and 10th pips; f_{other} gives the tone weights at all other positions. After the tone-weighting function is applied to the tone, the temporal-weighting function, *t,* gives the weight for each temporal position. We sum up all these products for the statistic θ. **C.** Our model has been simplified so that if θ + *Noise* is greater than the *Criterion*, then the participant says "Major." If this sum is less than the *Criterion*, then they say "Minor."

Parameter	Constraints	Degrees of Freedom
Token-Weighting Functions- $f_{final}, f_{accented}, f_{other}$ (13 parameters per function)	1. Mean = 0 2. Sum of Squares = 1	11 df (per function)
Temporal-Weighting Function- *t* (15 parameters)	none	15 df
Criterion Parameter (1 parameter)	none	1 df

Table 2.1 *Full Model Constraints and Degrees of Freedom for Extended-Notes and Rests Conditions.* Separate analyses were run for each participant in each condition. In the Extended-Notes and Rests Conditions, the full model had a total of 55 parameters with 49 df.

Non-rhythmic Condition. For the Non-Rhythmic Condition analysis, the full model differed in two ways:

1. There were only two token-weighting functions: f_{final} and f_{other}
2. We introduced 3 *Criterion* parameters

First, there were only two token-weighting functions because the Non-Rhythmic condition included no accented notes, so there was no purpose for the $f_{accented}$ function. f_{final} operated exactly the same as before, but f_{other} included all temporal positions (1:14) except for the final position (see **Equation 2.4 & 2.5**). Secondly, 3 *Criterion* parameters were used to manage the 4 possible responses in the Non-Rhythmic condition. **Equations 2.6-2.9** demonstrate how these criterion parameters mapped to different responses.

Let,

$$\theta = \sum_{k=1}^{15} f_k(tone_k) \ * t(k) \tag{2.4}$$

where f_k comprised two functions such that

$$f_k(tone_k) = \begin{cases} f_{final}(tone_k), & k = 15 \\ f_{other}(tone_k), & otherwise \end{cases} \tag{2.5}$$

In any given trial of the Non-Rhythmic Condition, we assumed that if

$$\theta + Noise < Criterion_1 \tag{2.6}$$

participants would respond "Very Minor." If

$$Criterion_1 < \ \theta + Noise < Criterion_2 \tag{2.7}$$

participants would respond "Minor." If

$$Criterion_2 < \ \theta + Noise < Criterion_3 \tag{2.8}$$

participants would respond "Major." If

$$\theta + Noise > Criterion_3 \tag{2.9}$$

participants would respond "Very Major."

The full model parameters and constraints for the Non-Rhythmic Condition are given in **Table 2.2** below.

Parameter	Constraints	Degrees of Freedom
Token-Weighting Functions- f_{final}, f_{other} (13 parameters per function)	1. Mean = 0 2. Sum of Squares = 1	11 df (per function)
Temporal-Weighting Function- t (15 parameters)	none	15 df
Criterion Parameters (3 parameters)	none	3 df

Table 2.2 *Full Model Constraints and Degrees of Freedom for the Non-Rhythmic Condition.* Separate analyses were run for each participant. In the Non-Rhythmic Condition, the full model had a total of 44 parameters with 40 df.

Model Fitting

The Full Model was fit with a separate Markov Chain Monte-Carlo (MCMC) sampling procedure for each participant in each condition. See **Appendix B** for details on the MCMC sampling procedure. 200,000 samples were collected for each model fit. All figures in the Results section display parameter means and 95% credible intervals from the last 90,000 samples.

Results

Model Comparisons

In order to assess the separability of pitch and rhythm, we compared various possible models using likelihood-ratio tests. The results are below in **Table 2.3**.

Extended-Notes Condition

Restricted	Full	Participant 1 X^2, *df*, *p-val*	Participant 2 X^2, *df*, *p-val*	Participant 3 X^2, *df*, *p-val*
Separable $f_{other} = f_{accented} = f_{final}$	*Full* f_{other}, $f_{accented}$, f_{final}	43.31, 22, .004	80.41, 22, < .001	68.42, 22, < .001
f_{other}, $f_{accented} = f_{final}$	*Full*	22.65, 11, .020	19.67, 11, .050	15.98, 11, .142
$f_{other} = f_{accented}$, f_{final}	*Full*	18.89, 11, .063	31.39, 11, .001	37.28, 11, < .001
$f_{other} = f_{final}$, $f_{accented}$	*Full*	25.94, 11, .007	62.51, 11, < .001	48.86, 11, < .001

Rests Condition

Restricted	Full	Participant 1 X^2, *df*, *p-val*	Participant 2 X^2, *df*, *p-val*	Participant 3 X^2, *df*, *p-val*
Separable $f_{other} = f_{accented} = f_{final}$	*Full* f_{other}, $f_{accented}$, f_{final}	38.91, 22, .015	70.13, 22, < .001	47.33, 22, .001
f_{other}, $f_{accented} = f_{final}$	*Full*	12.78, 11, .308	36.31, 11, < .001	15.44, 11, 0.16
$f_{other} = f_{accented}$, f_{final}	*Full*	6.28, 11, .854	15.92, 11, .144	10.98, 11, .445
$f_{other} = f_{final}$, $f_{accented}$	*Full*	35.19, 11, < .001	55.09, 11, < .001	37.65, 11, < .001

Non-Rhythmic Condition

Restricted	Full	Participant 1 X^2, *df*, *p-val*	Participant 2 X^2, *df*, *p-val*	Participant 3 X^2, *df*, *p-val*
Separable $f_{other} = f_{final}$	*Full* f_{other}, f_{final}	41.91, 11, < .001	86.68, 11, < .001	40.99, 11, < .001

Table 2.3 *Likelihood-Ratio Tests for Separability of Rhythm and Pitch.* In each condition, restricted models (less tone-weighting functions depending on temporal position) were compared to the full model. In all conditions for all participants, the *Full Model* fit the data significantly better than the *Separable Model* which only had a single tone-weighting function irrespective of temporal position. Highlighted cells contain p-values less than .05, so we can reject the hypothesis that these models are equally likely for the data.

Table 2.3 presents likelihood-ratio tests that reveal:

1. In all conditions for all participants, the *Full Model* provided a significantly better fit to the data than the *Separable Model,* which only had a single tone-weighting function.
2. In all conditions for all participants, the *Full Model* provided a significantly better fit to the data than the $f_{other} = f_{final}$ *Model*. Having separate functions for f_{other} and f_{final} is always necessary for the best fit.
3. In the Rests Condition, the $f_{other} = f_{accented}$ *Model* was not significantly different from the *Full Model*. The accented pips did not create significantly different tone-weights compared to the other pips in the Rests Condition.
4. The remaining model comparisons varied across participants, but it is apparent that the *Full Model* was superior in more instances for the Extended-Notes Condition versus the Rests Condition. This suggests that rhythm and pitch interacted less separably in the Extended-Notes Condition.

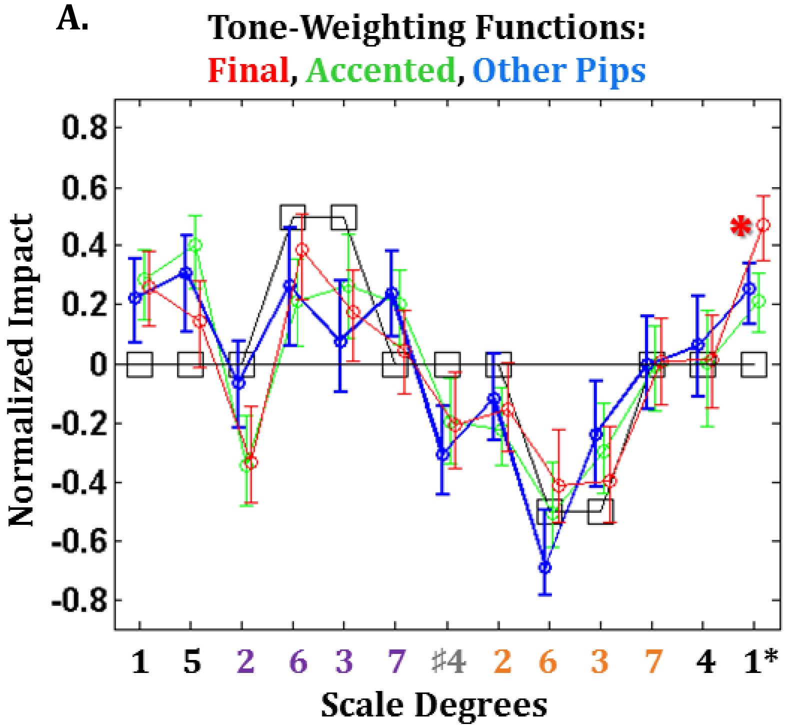
The following sections plot the results from the Full Model.

Extended-Notes Condition

Tone-Weighting Functions. The tone-weighting functions are plotted in **Figure 2.4**. Participants were influenced to respond "major" by the tones with positive impacts, and they were influenced to respond "minor" by the tones with negative impacts. In **Figure 2.4,** the Other Pips Function (f_{other} in **Equation 2.3**) gives the relative impact of tones that were not at the special rhythmic locations of the accented and final tones. The Accented Pips Function ($f_{accented}$ in **Equation 2.3**) gives the relative impact of the extended notes in the sequence at temporal positions 5 & 10. The Final Pip function (f_{final} in **Equation 2.3)**

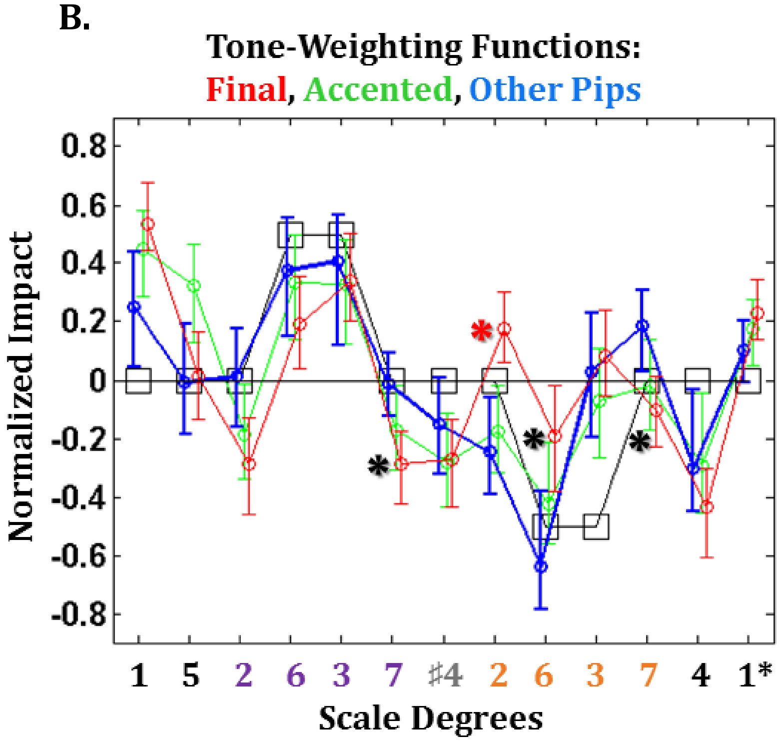
gives the relative impact of tones at the final pip, this was also an extended note. Any significant differences between these three functions point to an interaction of rhythm and pitch.

A.



* Final tone-weight CI different from Accented & Other tone-weights

B.



* Final tone-weight CI different from Accented & Other tone-weights

* Final tone-weight CI different from Other tone-weight

C.

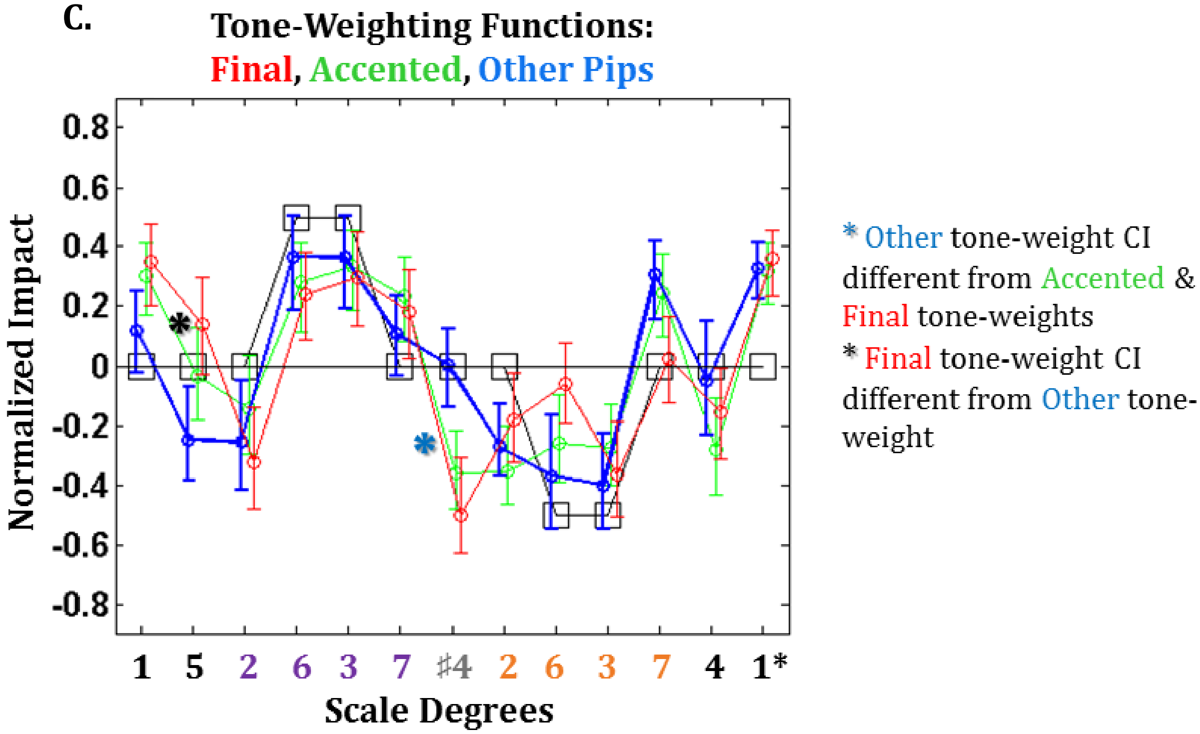


Figure 2.4 *Extended-Notes Condition: Tone-Weighting Functions*. All points are the means of MCMC samples which reflect a stable estimate of the posterior density. The error bars are 95% credible intervals. Positive impacts reflect a “major” and negative impacts reflect a “minor” influence on the decision statistic. The black line with squares shows the target function defined by the feedback rule. The major notes are labeled with the purple scale degrees and the minor notes are the orange scale degrees. The tone-weighting functions are plotted f_{other} (blue), $f_{accented}$ (green), f_{final} (red). Asterisks denote differences between the 95% credible intervals between tone-weighting functions. Results from each participant are plotted as follows: **A.** Participant 1, **B.** Participant 2, **C.** Participant 3.

Participant 1. In **Figure 2.4A,** each tone-weighting function estimated that the tonics (1 & 1*) gave a major weight. This deviated significantly from the target function. This result was unexpected because the tonics should not influence a “major” or “minor” decision. In other words, the tonics should have a zero weight. Since the tonics establish the tonal center, it appears as though a stronger center makes a tone-scramble sound more “major” or happy.

For Participant 1, the majority of the minor weight was driven by the minor 6th. The minor 3rd, tritone (♯4) and the major 2nd, also influenced Participant 1 to decide that a tone-scramble was "minor." The influence of the tritone was interesting because it had been referred to as "the devil in music" during the Middle Ages (Crane, 1976), so understandably it can have the connotation of "minor" or "sad." The tritone is also a particularly unstable note in the tonal hierarchy (Krumhansl & Cuddy, 2010). The influences of the major 2nd and tritone were unexpected because the feedback rule (target function) did not encourage a "minor" weighting of either of these tones.

Participant 1 had only one difference (assessed by 95% credible interval differences) between the three tone-weighting functions. The Final Pip Function gave a more "major" weight to the high tonic (1*) than the Accented and Other Note Functions. When the high tonic (1*) was the final tone in a tone-scramble, it had a more "major" impact compared to when the high tonic was at any other temporal position in the tone-scramble. This is evidence for an interaction between the influence of temporal position and tone-type. Ending on a tonic provides the strongest completeness/resolution ratings (Boltz, 1989), so this result suggests that resolution can make a tone-scramble sound more "major"/happy.

Participant 2. In **Figure 2.4B**, we see a similar pattern of a "major" influence for the low tonic (1) and the high tonic (1*). The "major" impact of the tonic is reliable for both the Accented and Final Pip Functions. Again, the majority of the "minor" influence came from the minor 6th. An unexpected deviation from the target function occurred at the perfect 4th (4). Participant 2 gave the perfect 4th a stronger minor impact than the minor 3rd.

For Participant 2, there were 4 differences between the tone-weighting functions. The first 3 differences were between the Final Pip Function and the Other Pips Function. These occurred at the major 7th, the minor 6th, and the minor 7th. The final difference was between the Final Pip Function and both the Accented Pips and the Other Pips Functions. This difference occurred at the minor 2nd. These differences were difficult to interpret except for the major 7th. Since the major 7th is the last chromatic tone before returning to the tonic, it is often called the "leading-tone," and ending on the leading-tone gives a very poor sense of completeness or resolution (Boltz, 1989). This lack of resolution might create a stronger "minor"/"sad" influence if resolution has an important influence on major/minor discrimination.

***Participant 3*.** Again, we see in the Accented and Final Pip Functions, the tonics (1 & 1*) are elevated above zero to have a "major" impact on the participant's decisions.

There were two differences between the tone-weighting functions. First, Participant 3 weighted the perfect 5th (5) significantly more "major" when it was the final pip versus the "other" (non-accented) pips. This result aligns with the notion that resolution improves the "major"/"happy" sound of a tone-scramble because the perfect 5th (a.k.a. the dominant) is a very stable tone to end on. Like the tonic, it is shared in both major and minor modes, so the standard music theory would suggest that it should not reinforce the discrimination between major/minor. Next, Participant 3 weighted the tritone (♯4) more "minor" at the Accented and Final Pip temporal positions versus the Other Pip positions. In other words, whenever the tritone was an extended note, it made a more "minor" contribution to Participant 3's decision statistic. The impact of this single note for this participant the only clear interaction between the non-final accented notes and pitch in our entire data set.

Summary. For all participants, there were some systematic deviations from the target function. This failure to match the target function reflects some processing limitations in attempting to match the feedback rule. For all participants, there were always positive weightings, or a "major" impact, for each of the tonics (1 & 1*) at the extended notes. Whenever the tonic was accented, the participant received some influence to respond "major." Since the tonic is the tonal center, it seems like a more centered tone-scramble sounds more "major"/happy.

The interactions between temporal position and tone type were different for each participant. Most of these interactions were found only for the Final Pip function which reflected the impact of the note of resolution. This supports the likelihood-ratio test results because the Final Pip Tone-Weighting Function needed to be distinct from the Other Pip Function for the best fit to the data. In short, in the Extended-Notes Condition, rhythm and pitch were not separable in the way they interacted in major/minor classification.

There was only one case, the tritone (♯4) for Participant 3, in which all the extended tones, in both the Accented and Final Pip functions, had a different impact than the non-accented tones. For this participant, the tritone gave a stronger "minor" impact when it was one of the special extended notes. This created the opposite effect of the tonic; the tritone is one of the least stable notes, and this instability appears to be related to the "minor" decision for Participant 3.

Temporal-Weighting Function and Criterion. The temporal functions (t in **Equation 2.3**) in **Figure 2.5** indicated large boosts of sensitivity to the accented pips at temporal positions 5, 10, and 15. All of these sensitivity increases reflect the separable manner in which extended notes amplify sensitivity to any particular tone at that temporal position. The final pip, at temporal position 15, had the strongest impact, especially for Participant 2. For participant 3, the first tone of the sequence also had a higher impact than many of the other non-accented tones. The overall systematicness of the following the tone-weighting functions is reflected by the height of the temporal function. Our participants performed quite systematically.

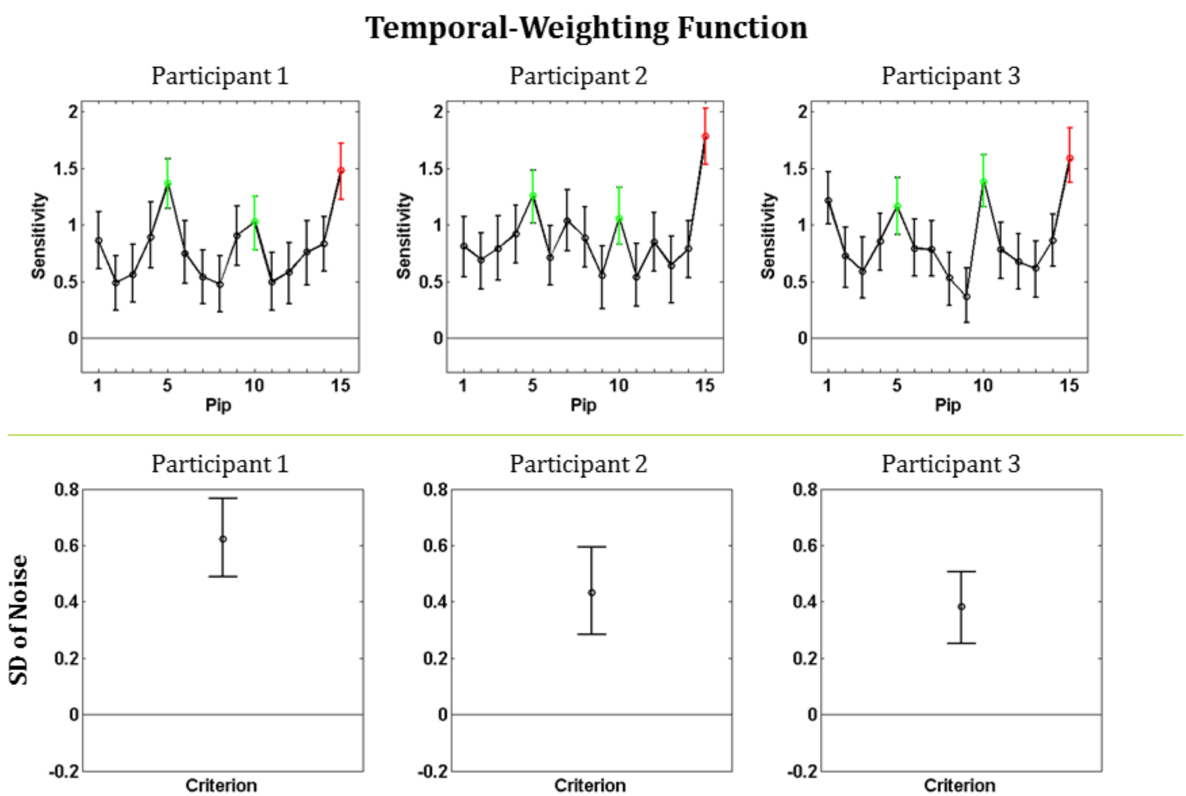
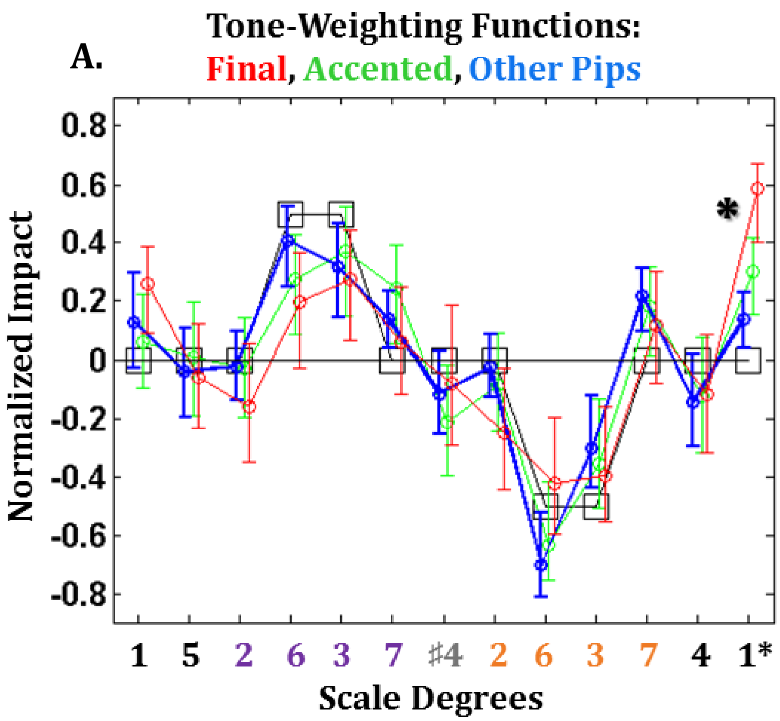


Figure 2.5 *Extended Notes Condition: Temporal-Weighting Functions and Criterion Parameters*. All points are the means of the MCMC sampling and error bars are 95% credible intervals. **Top Panel.** The extended notes at temporal position 5 & 10 are shown in green, and the extended note at the final position 15 is in red. For all participants, the estimated temporal-weighting function shows strong boosts at the extended notes (pips) in the stimulus. The final pip had equal or higher impact than the other accented notes. **Bottom Panel.** The criterion was unexpectedly positive for all participants.

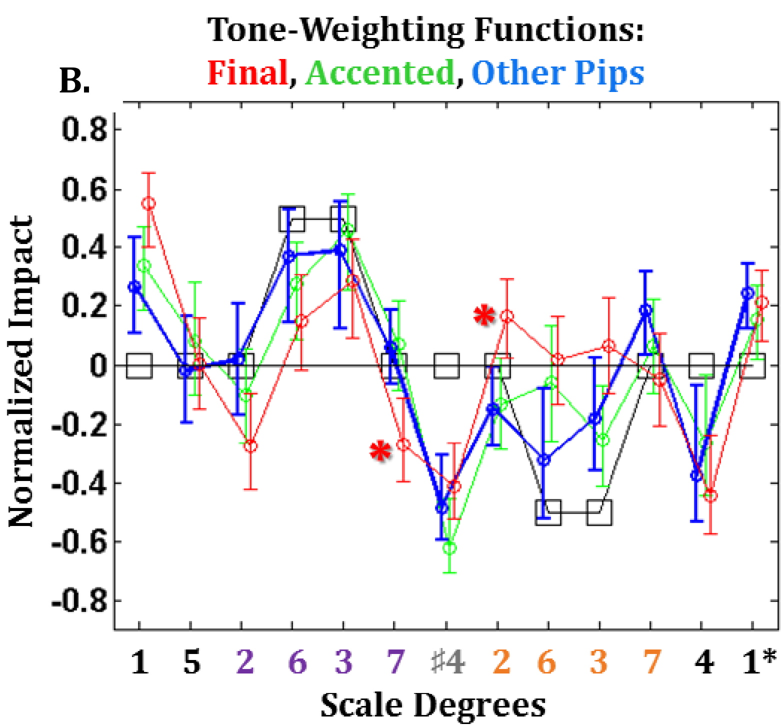
For all three participants, the criterion parameter was positive, indicating a bias to respond "major." This result was unexpected since all participants were highly sensitive to the major/minor task.

Rests Condition

Tone-Weighting Functions. The tone-weighting functions for the Rests condition are plotted in **Figure 2.6**. This figure uses the same plotting conventions as the Extended-Notes results plotted in **Figure 2.4.**



* Final tone-weight CI different from Other tone-weight



* Final tone-weight CI different from Accented & Other tone-weights

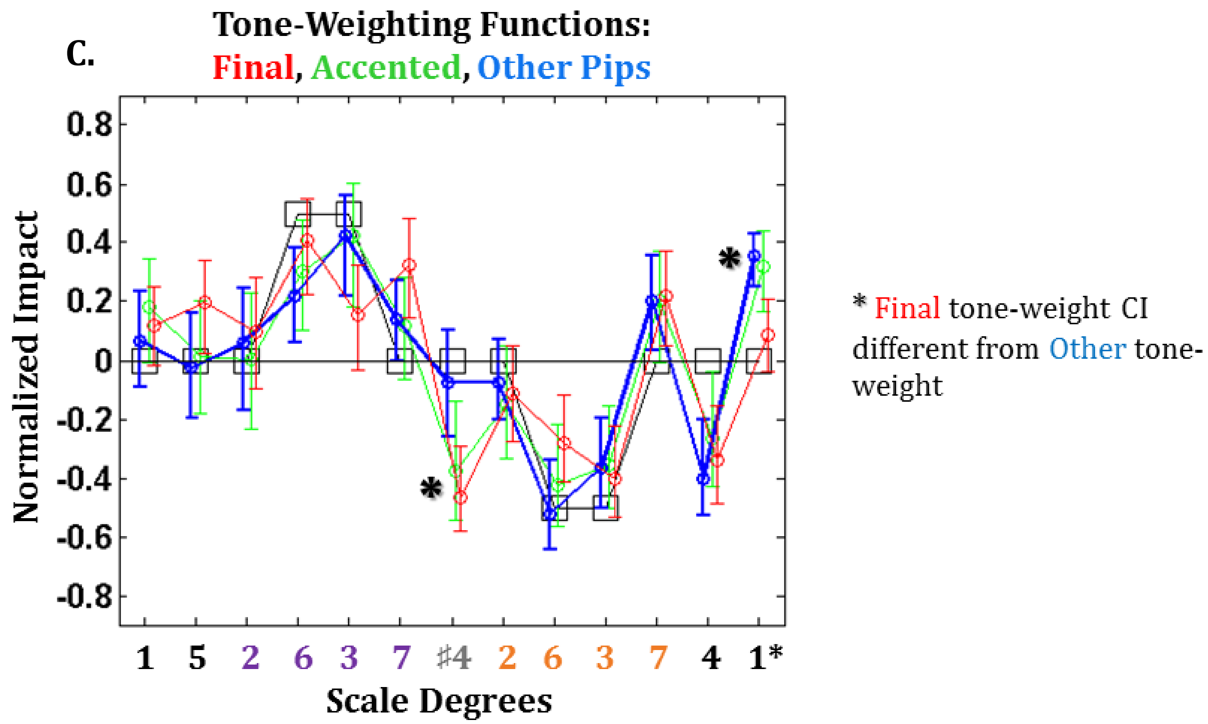


Figure 2.6 *Rests Condition: Tone-Weighting Functions*. All points are the means of MCMC samples which reflect a stable estimate of the posterior density. The error bars are 95% credible intervals. Positive impacts reflect a "major" and negative impacts reflect a "minor" influence on the decision statistic. The black line with squares shows the target function defined by the feedback rule. The major notes are labeled with the purple scale degrees and the minor notes are the orange scale degrees. The tone-weighting functions are plotted f_{other} (blue), $f_{accented}$ (green), f_{final} (red). Asterisks denote differences between the 95% credible intervals between tone-weighting functions. Results from each participant are plotted as follows: **A.** Participant 1, **B.** Participant 2, **C.** Participant 3.

Participant 1. In **Figure 2.6A,** Participant 1 was close in matching the target function for this condition. Similar to the Extended-Notes Condition, the final pip function had a positive weight for both tonics (1 & 1*). Again, the minor 6th had the strongest minor impact for this participant.

There was a difference between the Final Pip function and the Other Pips Function at the high tonic (1*). When tone-scrambles ended on the high tonic, 1*, it had a stronger “major” influence than when the high tonic was in non-accented temporal positions in the tone-scramble. This participant showed the same effect in the Extended-Notes condition.

Participant 2. As in the Extended-Notes Condition, both tonics (1 & 1*) had a major contribution to Participant 2’s decision statistic for the Rests Condition (**Figure 2.6B**). Again, perfect 4th (4) had a strong minor influence for this participant. Even though the 4th is modernly considered a “perfect” interval, it has been called a dissonance by Fux’s fundamental book on counterpoint, *Gradus Ad Parnassum* (Mann, 1965). This fact may have created a context for Participant 2 to experience the 4th as a “minor” tone. In this condition, the tritone (♯4) also had a powerful impact on this participant’s decision to classify a tone-scramble as “minor.”

There were two differences between tone-weighting functions, and for both, the Final Pip Function tone weight was different from both the Other and Accented Pips Functions. These two differences were found at the major 7th and the minor 2nd. The major 7th had a stronger “minor” impact and the minor 2nd had a stronger “major” impact when it was the final pip in a tone-scramble. This participant had differences that followed a similar pattern at both of these tones in the Extended-Notes Condition.

Participant 3. This participant achieved tone-weighting functions that more closely matched the target function compared to their results in the Extended-Notes Condition. The high tonic (1*) has a “major” impact and the perfect 4th (4) has a strong “minor” impact. The perfect 4th effect is similar to the one we observed in both conditions for Participant 2.

There were two tone-types where the Final Pip Function weight was different from the Other Pips Function weight. First of all, the tritone (♯4) had the extra “minor” impact when it was the final pip in a sequence. This matches this participant's results in the Extended-Notes Condition. Secondly, the high tonic (1*) had an interesting interaction. When the final pip was the high tonic, the sequence was *less* "major" sounding than when the high tonic was at the Other Pip (unaccented) temporal positions. This was a peculiar result specific to only this participant in this condition.

Summary. In the Rests Condition, the tonics (1 & 1*) did not always have a "major" impact on participant's decisions. Participants 1 & 3 were closer to achieving the target function compared to their results in the Extended-Notes Condition.

There were some differences between tone-weighting functions; these differences revealed interactions between tone-type and temporal position at the final pip. For each participant, these interactions were a subset of their interactions in the Extended-Notes Condition. Again, we observe that the Final Pip Tone-Weighting Function must be distinct from the Other Pip Function for the best fit to the data. In the Rests Condition, rhythm and pitch were not separable in the way they interacted in major/minor classification.

Temporal-Weighting Function and Criterion. The metrical accents created by the rests did not give boosts of sensitivity as strong as the extended-notes did in the first condition (**Figure 2.7**). Participant 1 showed very few differences in their sensitivity to temporal positions of pips in a tone-scramble. Participant 2 had the strongest rhythmic pattern in their temporal-weighting function, and they were particularly sensitive to the final pip. Participant 3 showed a nearly linear pattern of sensitivity; the participant became more and more sensitive to tones as the scramble progressed.

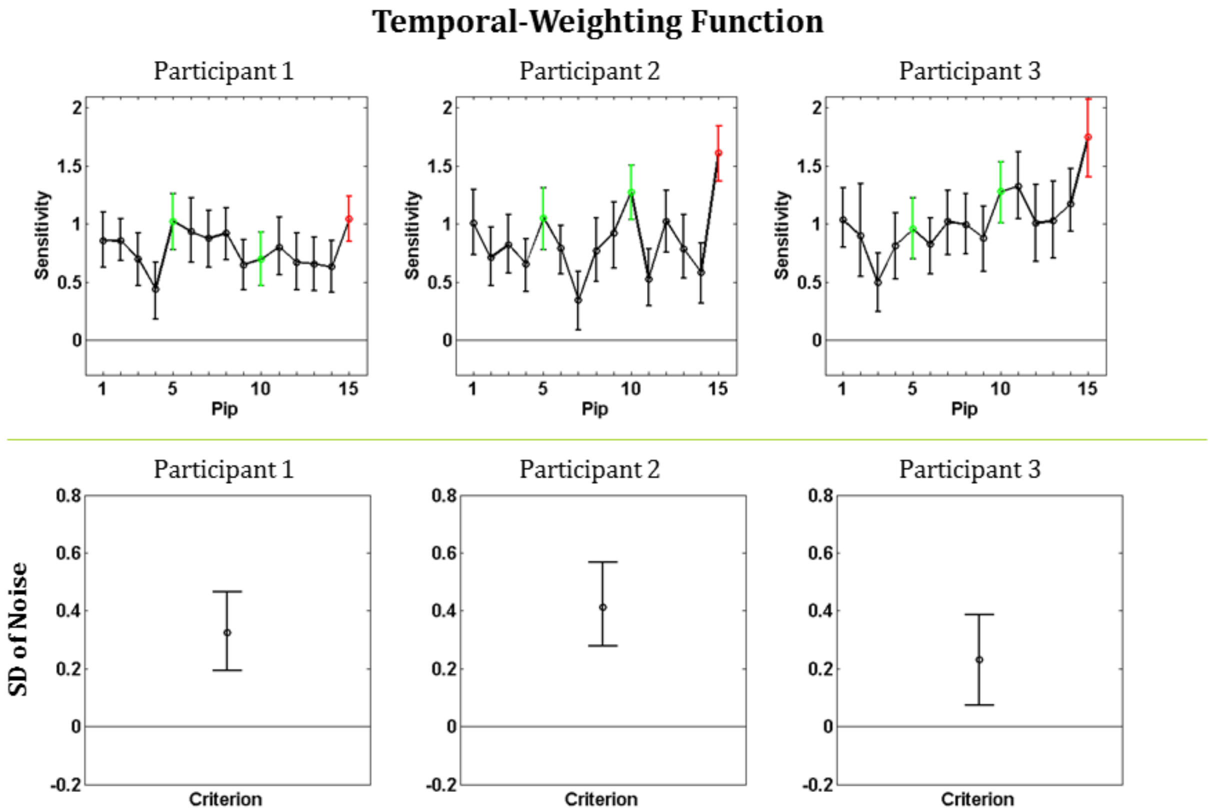
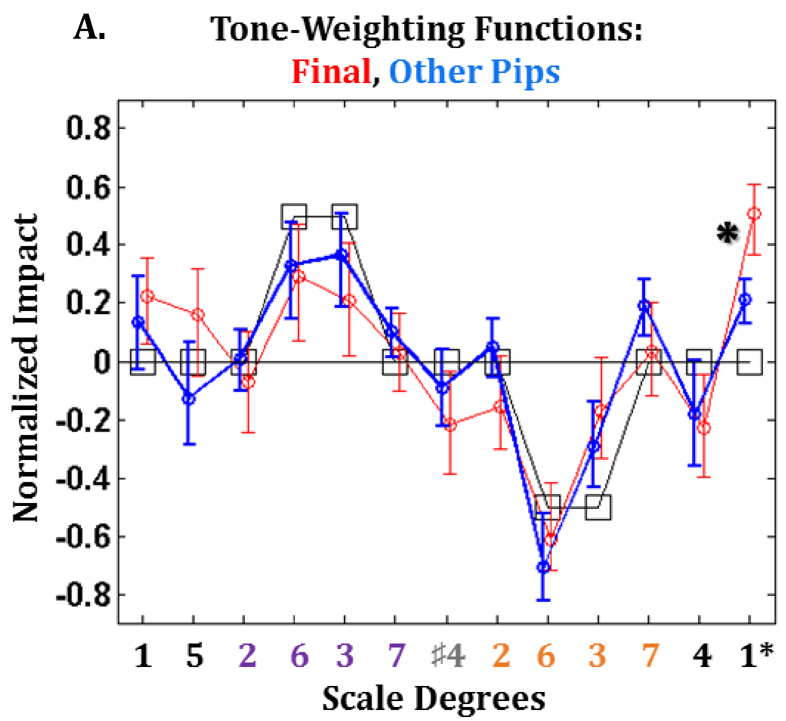


Figure 2.7 *Rests Condition: Temporal-Weighting Functions and Criterion Parameters.* All points are the means of the MCMC sampling and error bars are 95% credible intervals. **Top Panel.** The metrical accents at temporal position 5 & 10 are shown in green, and the metrical accent at the final position 15 is in red. The temporal weighting functions show an overall weaker additive effect of rhythmic accents compared to the Extended-Notes Condition. **Bottom Panel**. As in the previous condition, the criterion was positive for all participants.

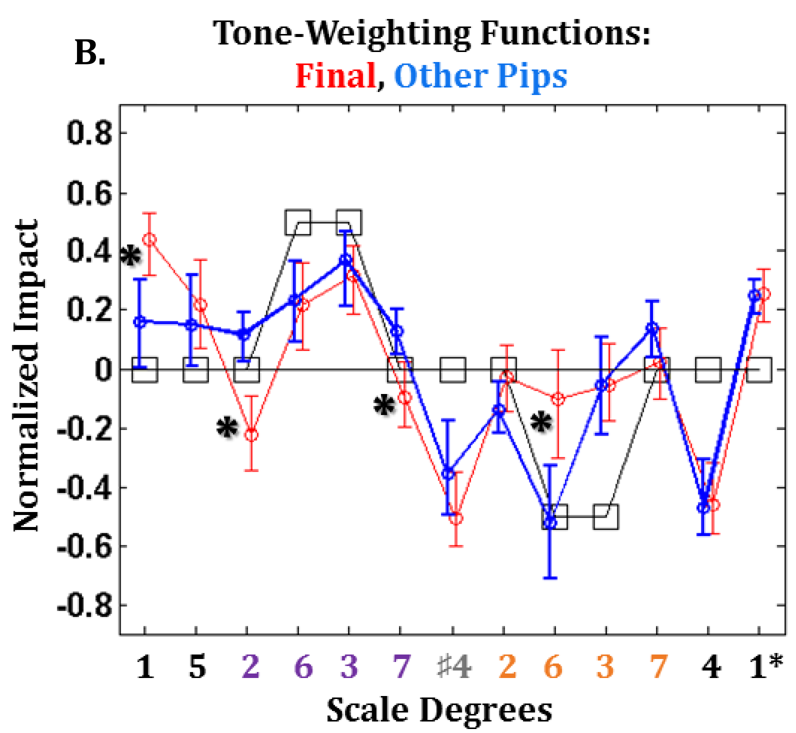
In **Figure 2.7,** the criterion parameter was positive for all participants, and this provides evidence for the "major" bias again.

Non-Rhythmic Condition

Tone-Weighting Functions. In **Figure 2.8**, the Non-Rhythmic Condition shows two tone-weighting functions: the Final Pip Function and the Other Pips Function because there were no rhythmically accented tones (see **Equation 2.5**).



* Final tone-weight CI different from Other tone-weight



* Final tone-weight CI different from Other tone-weight

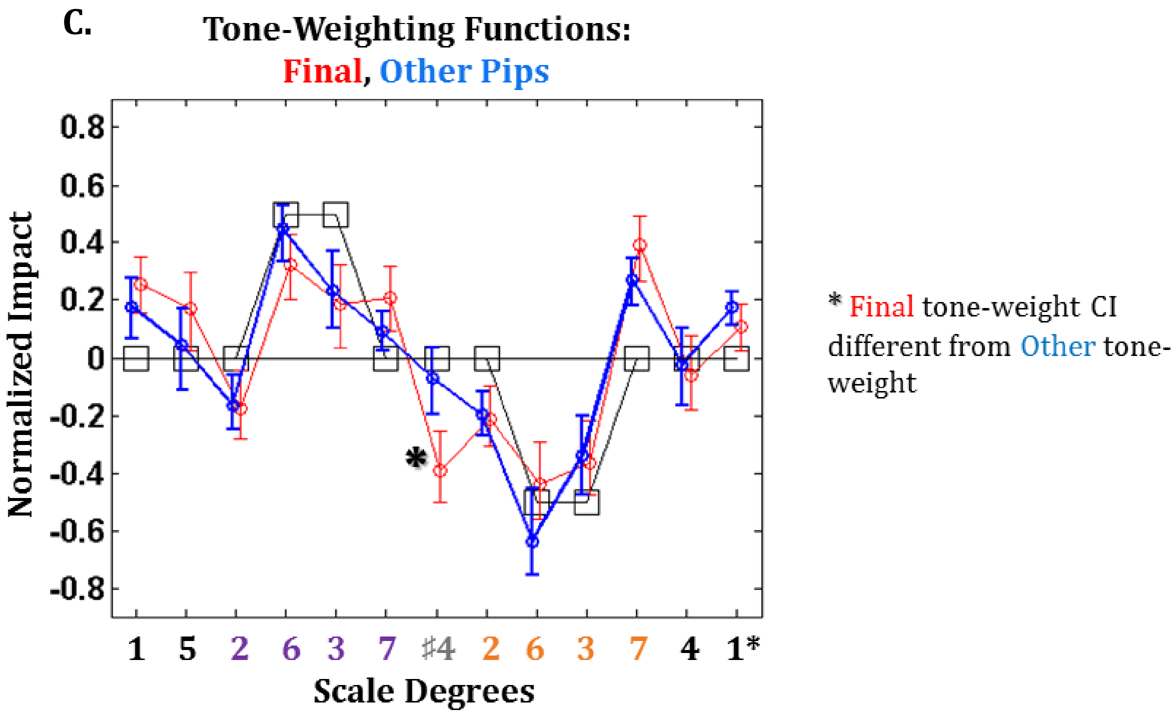


Figure 2.8 *Non-Rhythmic Condition: Tone-Weighting Functions*. All points are the means of MCMC samples which reflect a stable estimate of the posterior density. The error bars are 95% credible intervals. Positive impacts reflect a "major" and negative impacts reflect a "minor" influence on the decision statistic. The black line with squares shows the target function defined by the feedback rule. The major notes are labeled with the purple scale degrees and the minor notes are the orange scale degrees. The tone-weighting functions are plotted f_{other} (blue) and f_{final} (red). Asterisks denote differences between the 95% credible intervals between tone-weighting functions. Results from each participant are plotted as follows: **A.** Participant 1, **B.** Participant 2, **C.** Participant 3.

Participant 1. Participant 1 showed a few similar effects found in the previous conditions. To begin with, there was always a "major" impact of the tonics (1 & 1*) when they were the final pip in a tone-scramble. Next, the minor 6th had the strongest impact on the participant's decision to respond "minor." Lastly, the high tonic was weighted more "major" when it was the final pip versus all other temporal positions. This interaction was found in all 3 conditions for Participant 1.

Participant 2. The tonics (1 & 1*) always had a "major" impact on Participant 2's decision statistic. As in the other conditions, the perfect 4^{th} (4) had a strong "minor" impact for Participant 2. The minor 6^{th} and the tritone also had strong "minor" weights.

There were 4 differences between the tone-weighting functions. Three of these differences demonstrated unexpected tone-weights at the final pip. The major 2^{nd} and major 7^{th} had more "minor" impacts, and the minor 6^{th} had a *less* "minor" impact when they were at the final pip position compared to all other temporal positions. The major 7^{th} had this final pip interaction in all conditions for Participant 2. The last difference was the low tonic (1) had a higher "major" impact when it was at the final pip compared to other pip positions. The low tonic resolution had the same effect as the high tonic resolution- ending on any tonic gave a particularly complete and resolved feeling that can relate to a "major" classification in our paradigm.

Participant 3. Like Participant 2, the tonics (1 & 1*) always had a "major" impact on Participant 3's major/minor discrimination. The minor 6^{th} had the largest "minor" impact. The minor 7^{th} had a surprisingly high "major" impact; this effect was true to a lesser extent the previous conditions.

There was one interaction between pip position and tone-weight. Participant 3 weighted the tritone (♯4) more "minor" when it occurred at the final pip position compared to the other pip positions in a tone-scramble. This participant demonstrated this interaction in all conditions.

Summary. For all participants, the tonics (1 & 1*) had a "major" impact at the final temporal position. Each participant had at least one difference between tone-weighting functions even though the final pip was not structurally accented (e.g. in the Extended-

Notes Condition) or metrically accented (e.g. in the Rests Condition). These differences represent interactions between tone-type and temporal position, and these interactions were different between participants but showed some consistency across conditions for each participant. Even without rhythmic information, the *Full Model* was crucial to distinguish the important differences of tone-impacts at the final pip position.

Temporal-Weighting Function and Criterion. The temporal-weighting functions were relatively flat except for strong boosts of sensitivity at the final pip for Participants 2 and 3 (**Figure 2.9**). These boosts at the end demonstrate that there was a natural importance to the final pip even when there were no rhythmic accents in the stimuli. The final note is crucial to the notion of musical resolution, so resolution seems to play an important role in our participants' major/minor discrimination.

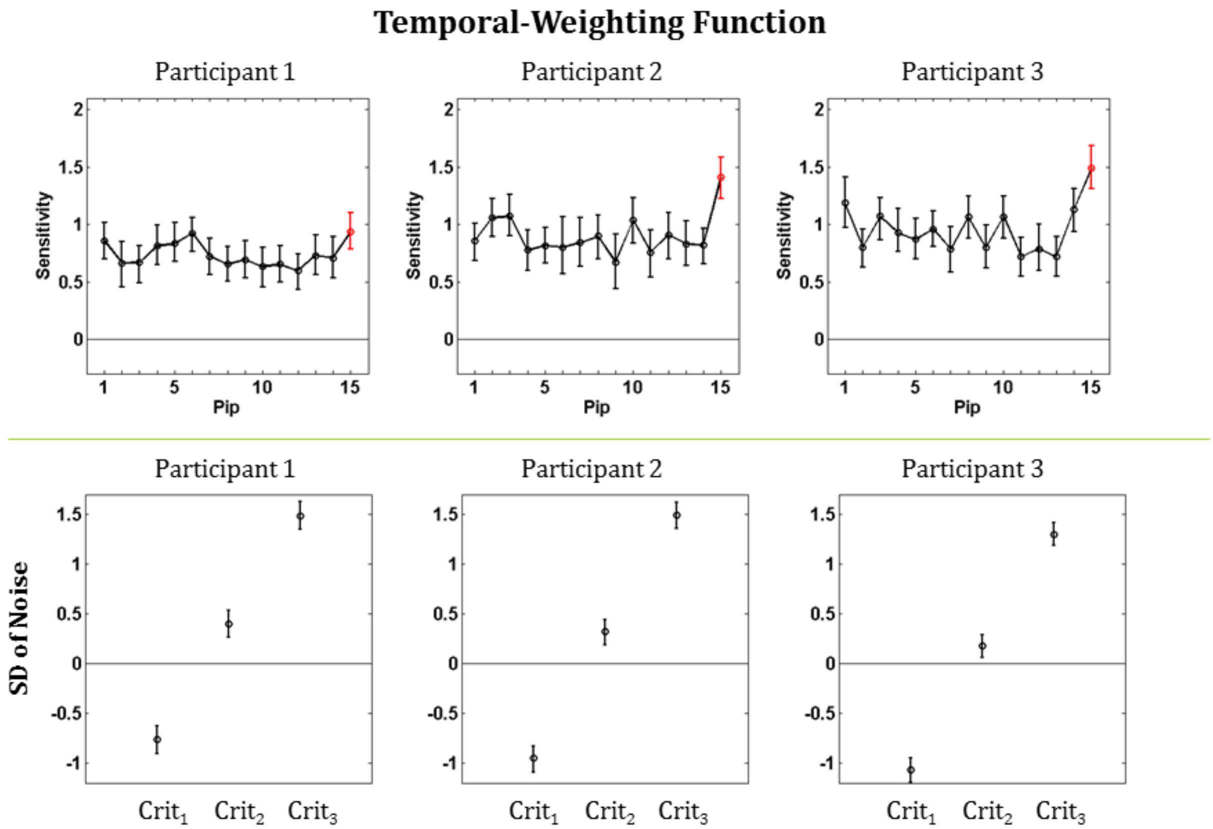


Figure 2.9 *Non-Rhythmic Condition: Temporal-Weighting Functions and Criterion Parameters.* All points are the means of the MCMC sampling and error bars are 95% credible intervals. **Top Panel.** The final pip at temporal position 15 is in red. There are significant boosts in sensitivity to the final pip for Participants 2 & 3. **Bottom Panel.** For all participants, the three *Criterion* parameters are evenly spaced with a positive ("major") bias.

The model for the Non-Rhythmic Condition (**Equation 2.6-2.9)** included 3 *Criterion* parameters. All participants show systematic 1.25 standard deviation intervals between each *Criterion* parameter.

Discussion

When making major/minor classifications, rhythmic accents amplified sensitivity to tones. The phenomenal accents in the Extended-Notes Condition created a stronger sensitivity to accented tones than the metrical accents in the Rests Condition. This difference was not unusual because the extended tones were twice as long as the accented tones in the Rests Condition.

The rhythmic manipulations (phenomenal and metrical accents) did not consistently have an interactive effect with pitch. However, at the final tone in a tone scramble, there were significant interactions between the temporal position and pitch in making the major/minor classification. This indicated that *resolution* had a unique effect on the impact of particular tones in the major/minor discrimination. These interactions occurred in all conditions, including the Non-Rhythmic Condition; thus, the internal rhythmic patterns in the Extended-Notes and Rests Condition were not necessary to create a special influence of the final pip. Since an entire stimulus could be considered a musical phrase, the final pip could have been interpreted as a global rhythmic accent even though this rhythm is less proper than the local phenomenal and metrical accents. These results, along with likelihood ratio tests, proved that we must reject the separable model of rhythm and pitch where rhythmic accentuation only amplifies sensitivity to pitch.

Tonal stability based on tonal hierarchy (see Krumhansl, 2000 for a review) gave some explanations for the interactions that occurred at the final pip. There were individual differences in the final pip interactions, but each participant showed relatively consistent interactions across conditions. Participant 1 always weighted the final high tonic as extra "major." The tonic is the most stable tone in the tonal hierarchy, and this stability in the resolution could sound more "major"/happy. Participant 2 always gave the major 7th a "minor" boost when it occurred on the final tone – the major 7th is quite unstable, so this poor resolution appears to be "minor"/sad sounding. Participant 3 always weighted the final tritone, #4th, extra "minor" compared to tritones at other temporal positions. The extra "minor" impact from the tritone may also be related to stability of resolution since is one of the most unstable notes. Aside from these differences, a low or high tonic at the final pip always gave some influence for a participant to classify a scramble as "major." This was the case in all conditions and participants except for Participant 3 in the Rests Condition, but the means of their tonic tone-weight estimates were both above zero.

Each tone-type, or scale degree, impacted participants in different ways. The major and minor 3rds and 6ths certainly played a significant role for all participants, but many other tones also gave significant contributions towards their decisions. There were always systematic departures from the target functions. Deviations from the target function pointed towards processing limitations. Even with trial by trial feedback, participants were not able to follow the feedback rule, nor did they comply with the music theory definitions major/minor tones in the scale. In an unpublished experiment from the Charles Chubb lab (related to Chubb, 2013), participants were able to follow the same target function much more accurately when tone scrambles were more brief, 65 ms, and included more tones, 26 pips. These preliminary results suggest that the longer tones of the present experiment create more unexpected tone-weighting functions.

The *Criterion* parameter was positive for all participants in all conditions. This suggests a "major" bias. In support of a "major" bias, an ERP study found that musicians assumed sequences of tones were major until they heard a minor note (Halpern, Martin, Reed, 2008). There is another possible explanation for the positive *Criterion* estimates. All of the stimuli had more tonics relative to the other tones, and since they typically had a major influence, participants needed to adjust their criterion to compensate. The positive criterion did not result in an uneven distribution of major/minor responses.

An additional post-hoc analysis tested the effect of the last two notes being ascending or descending. Neither ascending nor descending patterns of the last two notes created a significant effect on the major/minor classification.

One potential criticism would be that participants were not explicitly instructed to ignore rhythm. One recent experiment suggests that explicit instructions to ignore rhythmic variations in a pitch judgment did not aid participants in successfully ignoring rhythm (Jones, Johnston, Puente, 2006).

What does this tell us about the “happy” and “sad” qualities of major and minor tonalities respectively? Our results may suggest that when music ends in a stable and centered manner (i.e. resolving to the tonic), it gets happier. Conversely, perceived sadness may come from unstable resolution.

Conclusions

When classifying brief sequences of tones as “major” or “minor”, rhythm amplified sensitivity to pitch except for interactive effects at the final tone (pip). These interactions supported an inseparable model of rhythm and pitch for major/minor discriminations. The final pip had its own special temporal importance of resolution. This global marker of rhythm was more powerful than the local rhythmic accents (extended-notes and tones before rests) within a sequence of tones. Phenomenal accents created stronger boosts of sensitivity than metrical accents. Tonics had an unexpected “major” impact on participants’ decisions and there were individual differences in the way different tones impacted their major/minor distinction. The participants did not strictly follow the music theory nor the experimental feedback rule for weighting the impact of each tone-type.

CHAPTER 3: Binding brightness and loudness: the limits of attention and the temporal window

Abstract

In the present study, we explored the role of attention in binding brightness and loudness. Attentional binding of loudness and brightness can only be achieved for inter-stimulus-intervals (ISIs) greater than a critical ISI called the temporal window of binding. The present study used new experimental and analytical techniques to investigate the limits of attention and the temporal window of binding brightness and loudness. On each trial, the observer viewed a quick stream of 18 gray disks, each accompanied by a simultaneous burst of auditory white noise. Three levels of disk brightness and of noise loudness created nine possible pairings or token-types. For each condition, participants attended to a particular feature of the stimulus stream: brightness-only, loudness-only, correlatedness of brightness and loudness, zero intensity (blank/silent) tokens, or maximum intensity tokens. Participants judged whether the attended feature was "high" vs. "low" for each stimulus stream. Using a linear process model, attention filters were derived based how much each token-type impacted the participants' judgments. The model estimated the probability of binding non-simultaneous disks and noises to derive the temporal binding window. We found converging evidence for the previously discovered limit of the multimodal temporal binding window. A hierarchical Bayesian analysis established evidence for four dimensions of brightness and loudness sensitivity across participants. Participants were able to attend to loudness only and ignore variations in brightness, but they had more trouble attending to brightness only and ignoring loudness. The finding that participants achieved different attention filters in different conditions

demonstrates that top-down attention can powerfully modulate the binding of loudness and brightness in dynamic displays.

We are constantly perceiving light and sound, and we can segregate events and objects depending on the way these percepts combine. Sometimes we can focus our attention on just our vision or hearing, but we often require silence while we read and close our eyes while we focus on music. These attempts to block out one modality in order to focus on the other suggest that cross-modal information combines without complete attentional control. This begs the question, how effectively can we use our attention to control the way we perceive combinations of light and sound? This project focuses specifically on the attentional control we can exert in segregating/binding brightness and loudness information in dynamic displays.

Multisensory Integration & Attention

Some research has found that people tend to rate brightness more highly when a light is paired with a loud sound than a light without a sound, but the increased brightness ratings have been explained by response biases rather than low-level sensory interactions. For example, Stein et al. (1996) reported that uninformative auditory noise increased brightness ratings; however, Odgaard (2003) established that this enhancement resulted from a response bias by manipulating the proportion of trials in which noise was paired with brightness. Ben-Artzi & Marks (1995) also proposed a response bias model for this effect; the model described participants who would lower their criteria for classifying a light as "bright" if a loud tone was presented simultaneously. In an unspeeded sensory discrimination task by Marks et. al in 2003, combinations of correlated levels of brightness and loudness led to increased accuracy driven by response biases as well. These findings demonstrate that there is more than just low-level multisensory integration, but they do not quite explain how attention modulates brightness and loudness binding.

There are different explanations for how multisensory integration and multisensory attention interact. The ventriloquism effect, where visual biases affect sound localization, has been explained as multisensory integration that occurs independently of both endogenous and exogenous spatial attention (Vroomen, Bertelson, De Gelder, 2001). On the contrary, ERP studies by Talsma and Woldorff (2005 & 2007) show that multisensory attention can affect multisensory integration. The present study will investigate the interaction of multisensory integration and attention by on exploring the extent to which selective attention can mediate our perception of brightness and loudness.

The Temporal Binding Window

The present study introduced a task in which quick bursts of sounds and lights needed to be parsed and integrated over time. We must consider the temporal limit of processing for this parsing. For a brightness and loudness judgment, the temporal binding window (TBW) is the amount of time you need to process combined information from a sound and light. TBWs have often been characterized by the probability of multisensory illusions with various levels of stimulus onset asynchrony (Stevenson, 2012). For example, in the sound-induced flash illusion, two beeps are paired with one flash of light, but it gives an illusion of two flashes (Shams, Kamitani & Shimojo, 2000). In this case, the TBW can be measured by the temporal gap between the flashes and beeps under which the illusion is still perceived. If the illusion is perceived with shorter offsets, then this would reflect a narrower temporal binding window.

Fujisaki and Nishida (2010) used different method of extracting a temporal binding window for multimodal stimuli. Fujisaki & Nishida argue for a central TBW that is insensitive to different peripheral sensory processing speeds from vision, hearing, and touch. In one of their audiovisual binding experiments, a stimulus consisted of a visual patch that oscillated between two equiluminant colors, C_1 and C_2 of different hues, and a simultaneous sound that oscillated between two pitches, P_1 and P_2 of equal loudness. Participants judged whether C_1 was aligned in time with the occurrences of P_1 or P_2. In other words, the participants judged whether the sequence was "in-phase" our "out-of-phase." The stimuli were presented at different rates of oscillation, and classification performance was measured. For participants to achieve at least 75% performance, the stimuli could not oscillate faster than 2-3Hz. This 2-3Hz limit translates to a 167-250ms temporal binding window. These stimuli were carefully designed to be constant in brightness and loudness, so participants were forced to use top-down attention to achieve the color-to-pitch binding required for making the judgments. It should be noted, however, that Fujisaki & Nishida experimented with changing the auditory component from pitch to loudness and there was no effect on the 2-3Hz binding limit.

The Present Study

The present study explored whether or not this universal temporal binding window would hold with a different type of stimulus, task, and TBW calculation. Our stimuli were composed of a rapid stream of simultaneous disks of varying brightness and noise-bursts of varying loudness. Our stimuli differed from the Fujisaki and Nishida 2010 experiment because the Fujisaki and Nishida stimuli used only two levels of each of the visual and auditory features, conforming to a fixed, oscillating pattern of the phase-judgment stimuli. Our stimuli was composed of particular amounts of disk brightness and noise loudness, but these disks and noise-bursts were presented in random order. These random variations created temporally broadband frequency content in both brightness and in loudness. In addition, the Fujisaki and Nishida experiment did not examine brightness and loudness binding, and since these features both mark the intensity of their respective sensory modalities, there may exist special processes available to bind them. Such processes might enable strategic binding options precluded by the stimuli of Fujisaki and Nishida (2010).

Instead of a phase judgment task, our experiment required participants (in different, separately blocked conditions) to selectively attend to particular audiovisual aspects of each stimulus. The first two conditions asking participants to either judge mean brightness-only or mean loudness-only in the stream of disks and noise-bursts. These conditions tested the degree to which participants could decouple brightness information from simultaneous loudness information and vice versa. Three more conditions tested binding abilities by asking subjects to judge relative amounts of particular combinations of brightness and loudness. The extent to which participants could attend to the features required in our experiment will reveal how multisensory integration and attention interact for brightness and loudness in temporally varying sequences.

The current experiment extracted a new sort of temporal binding window. Our stimuli lasted approximately 1.5 seconds, and audiovisual tokens were attended to within a continuous stream, providing the chance for the fine structure of the stimulus to blur together. We characterized this blur by estimating a TBW from our data that was defined by the probability of binding non-simultaneous brightnesses and loudnesses. Participants would bind non-simultaneous brightnesses and loudnesses if their TBW subsumed multiple brightness-loudness tokens. We extracted this temporal binding window to test the limit established by Fujisaki & Nishida (2010).

In the present study, the data analyses used linear process models and, in doing so, deviated from the typical analysis methods in the multisensory binding literature. The model was fit using Bayesian Markov Chain Monte Carlo sampling methods. We also fit the data using a hierarchical Bayesian model to discover the limits of brightness/loudness binding across conditions and participants.

Our selective attention tasks allowed us to explore the limits of attention on brightness and loudness binding. We expected to find a smaller TBW than the Fujisaki and Nishida limit because brightness and loudness are both on the intensity dimension light and sound, and the random variations in the stimulus (as opposed to strict oscillations) may enhance temporal processing speeds.

Method

Participants

Three individuals participated in this experiment (2 males and 1 female). Two of the participants were the authors, and the other was a paid undergraduate volunteer. Each participant had normal or corrected-to-normal vision and normal hearing. All participants gave informed written consent approved by the Institutional Review Board at the University of California, Irvine.

Apparatus & Stimuli

The experiment was run on MATLAB on an Apple iMac. The display screen had a resolution of 1920x1080 pixels, with a refresh rate of 60 Hz. Sounds were presented over Sennheiser HD 201 headphones.

For each trial, a quick stream of 18 gray disks (83 ms per disk) was presented in the center of the screen (see **Figure 3.1**). Each disk was 3.6cm in diameter with had a raised cosine envelope. The background was uniform grey. Each disk was paired with a simultaneous burst of auditory white noise of varying amplitudes. Each noise-burst had an amplitude rise and decay window of ~8 ms at the onset and offset to avoid clicks.

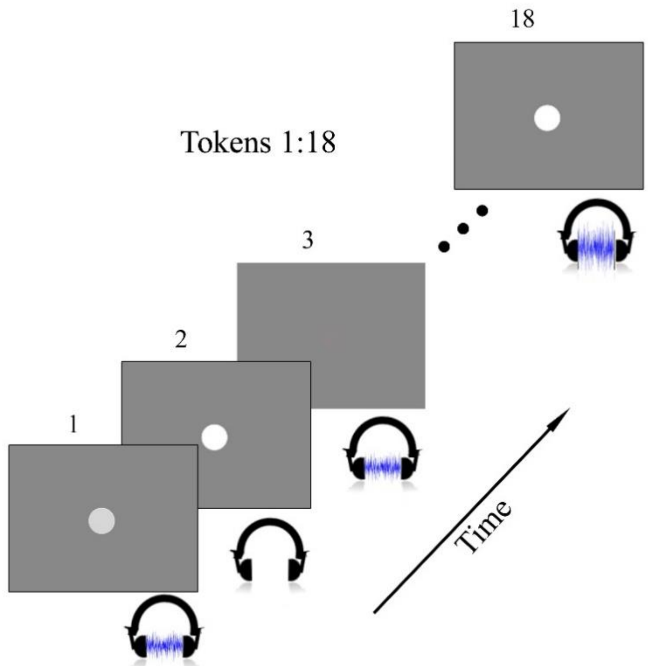


Figure 3.1 *Stimulus Presentation.* A sample stimulus of 18 simultaneous circles and noise-bursts that vary in amplitude.

Three levels of disk brightness and of noise loudness were used to produce 9 different token types of audiovisual pairings. The lowest brightness and loudness levels were completely blank (equivalent to the background grey) and silent (ambient room sound).

Brightness and loudness levels were chosen so that the middle amplitude appeared approximately half as intense as the high amplitude. Luminance measurements were taken with a PR-670 SpectraScan Spectroradiometer: blank (background grey) level = 25.51 cd/m^2, medium brightness level = 53.15 cd/m^2, high brightness level = 95.53 cd/m^2. Sound pressure level measurements were done with Brüel and Kjær 2260 Investigator. We used the A-weighted broadband detector with the fast time weighting setting for the following measurements: silent level = 33 dB SPL, medium loudness level = 69 dB SPL, high loudness level = 82 dB SPL.

Design

This experiment was a within-subjects design with 5 selective attention conditions. In the first two conditions, the participant was required to judge (1) mean brightness only and (2) mean loudness only of each stimulus stream. For example, in the brightness-only condition, the participant viewed and listened to a quick stream of audiovisual tokens and determined whether the mean brightness across all tokens was greater than a remembered standard established in previous trials. This judgment required the participant to ignore the irrelevant stimulus feature of loudness. The correlatedness condition (3) asked participants to determine whether the stimulus stream had more tokens that were correlated in their brightness/loudness intensity versus anti-correlated tokens. The last two conditions examined observers' abilities to selectively attend to relative values of just a single token type. The zero intensity condition (4) asked participants to make judgments based on the relative amount of only the zero intensity (blank/silent) tokens, and the maximum intensity condition (5) required participants to make judgments based on the relative amount of only the maximum intensity tokens (more precise definitions of each condition are to follow in **Figure 3.3**).

For every trial, a histogram of the 9 token types defined the 18 tokens that would be presented. These 18 tokens were presented in a random order for each trial. **Figure 3.2** shows how the *target filter* was used to give trial by trial feedback in that particular condition. Feedback was derived by applying the target rule to all the tokens in a stimulus and then adding up the values. The sign of this sum determined the correct response; in other words, if the sum was negative, the correct response would be “low” and if the sum was positive, the correct response would be “high.”

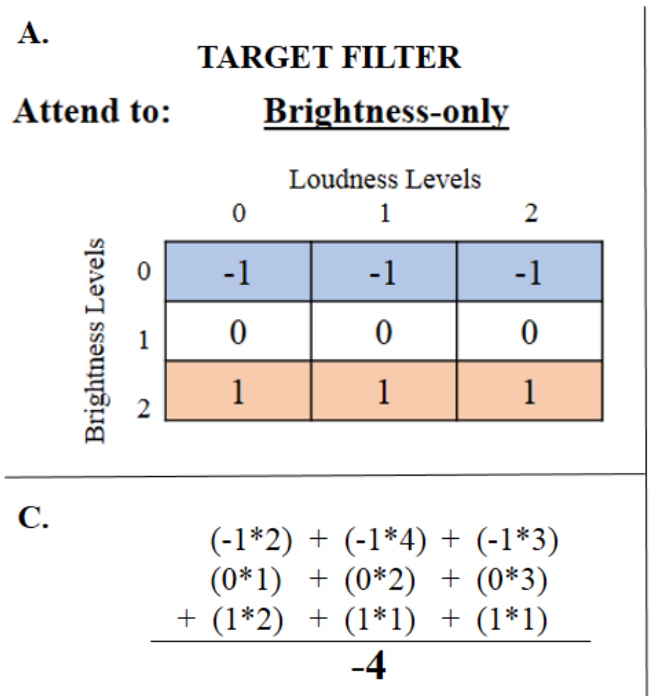


Figure 3.2 *Target Filter Example.* The target filter dictated the way participants should weight each token depending on the condition. **A.** For the brightness-only condition, the target filter weighted the low brightness tokens (blue) negatively and the high brightness tokens (red) positively. **B.** This is a sample stimulus-histogram defining the distribution of the 18 total tokens in a particular stimulus. **C.** When the target filter is applied to the stimulus, the negative sum indicates it is a “low” stimulus. If the participant responds “low”, the feedback will display “Correct.”

Figure 3.3 illustrates how each target filter precisely defined the task for each selective attention condition. Every target filter weighted some token-types positively and others negatively, and the weights in a target rule summed to zero.

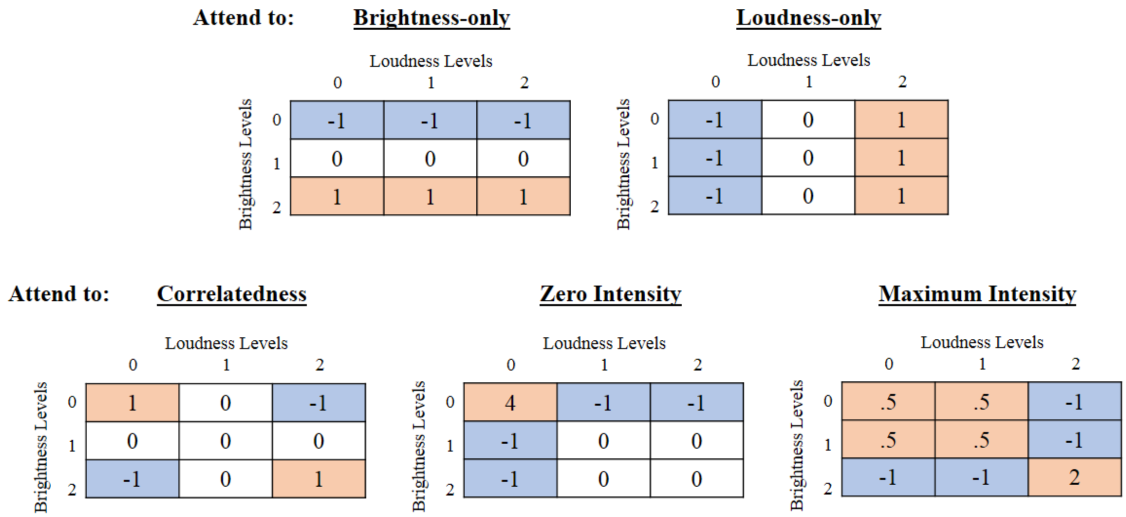


Figure 3.3 *Target Filters for Each Condition*. These target filter matrices show the positive (red) and negative (blue) weighting of each token type depending on the attention condition. When the target filter was applied to each stimulus, the sign of the sum determined whether the stimulus was "high" or "low" and correctness feedback was given accordingly.

In the brightness only and loudness only conditions, the stimulus-histograms were regulated by two, randomly interleaved 3-up-1-down staircases. One staircase controlled the histograms used in trials in which the correct response was "high" (i.e., trial in which the target function applied to the stimulus would yield a positive value); the other staircase controlled the histogram used in trials in which the correct response was "low." The "high" histogram began with many high-intensity brightness tokens (brightness-only condition) or high-intensity loudness tokens (loudness-only condition). The "low" histogram began with many low-intensity tokens for either brightness or loudness depending on the condition. In a given one of these staircases, if the participant responded correctly to the last three trials, then the stimulus-histogram would get more difficult (i.e., the histogram would be adjusted so that applying the target filter to the stimulus would yield a value closer to zero); otherwise, the stimulus-histogram would get easier. These adjustments included random variations across the various tokens that would adjust the mean brightness (brightness-only condition) or mean loudness (loudness-only condition). These 3-up-1-down staircases lead to approximately 79% correct performance in these tasks.

For the correlatedness condition, there was no staircasing procedure. Instead, the number of target tokens, either the correlated or anti-correlated tokens, was set by the experimenter to yield approximately 80% correct for each block of trials. For participants to achieve approximately 80% correct, the number of target tokens was set to 10 out of the 18 total tokens per stimulus. The remaining 8 tokens were chosen randomly.

For the zero intensity and maximum intensity conditions, the stimuli were created by using their target filter in **Figure 3.3** as orthogonal basis vectors to create positive and negative stimuli. In these conditions, the weights assigned to the non-emphasized tokens were intended to discourage strategies based on attending to brightness alone or loudness alone. The values in the maximum intensity target filter were further adjusted to better eliminate potential cues created by the unattended tokens. For each of these two conditions, 14 histograms were generated for each high and low stimulus type. Each trial randomly selected one of these histograms and then the tokens in that histogram were presented in random order. For the zero intensity condition, 2 medium brightness disks paired with medium loudness noise-bursts were presented at the start and end of each stimulus stream (22 total tokens). This was added to this condition so the participant could distinguish the zero intensity tokens from the start and end of the stimulus stream.

Procedure

The participant sat approximately 46cm from the screen with the lights off in the room. After each trial of 18 simultaneous disks and noise-bursts was presented, the participant responded with either a left- or right-arrow key-press; these key-presses corresponded to either low or high amounts of the attended feature for that condition. After each trial, feedback ("Correct" or "Incorrect") was displayed on the screen. The feedback in each condition encouraged participants to optimize their strategy based on the target filter.

Every participant ran the experiment in the following order of conditions: brightness-only, loudness-only, correlatedness, zero intensity, and maximum intensity. This order was maintained across participants to standardize training. After 100 trials of practice in each condition, testing was done in blocks of 100 trials. 1100 trials were analyzed for each participant in each condition.

Modeling

Basic Model

For each participant in each condition, a probit model (**Equation 3.1**) was used to measure the impact exerted on the observer's judgments by each of the 9 token types. The *attention filter,* $\vec{F}$, achieved by the participant in a given condition was composed of 9 parameters that reflected the relative weight exerted by each token-type on the internally computed statistic used by the participant to make his/her decision on each trial. The attention filter, $\vec{F}$ was constrained (1) to sum to zero and (2) to have a sum of squared values equal to 1. Thus, the parameter C in Eq. (3.1) reflected the consistency (relative to internal noise) of the participant in using his/her attention filter. The *Criterion* parameter reflected response bias, and the *Noise* in equation 3.1 was assumed to be a standard normal random variable. See **Table 3.1** for the full model constraints and degrees of freedom.

In any given trial of a particular condition, we assumed that the subject responded "low" (vs "high" otherwise) if

$$\sum_{k=1}^{18} F(token_k) \; * C + Noise < Criterion \tag{3.1}$$

where $token_k$ was the loudness-brightness pair occurring in position k (from 1 to 18). An $\vec{F}$ vector, C parameter, and *Criterion* parameter were estimated for each participant in each condition.

Full Model

Temporal Binding Parameters. The full model was developed with the addition of temporal binding parameters. These parameters revealed how much the simultaneous binding disks and noise-bursts contribute to the weighted sum reflecting the participant's decision. The addition of these parameters accounted for the possibility of “misbindings”: perceptual bindings of non-simultaneous brightnesses and loudnesses in the stimulus. See **Figure 3.4** for the possible perceptual bindings considered by the Full Model.

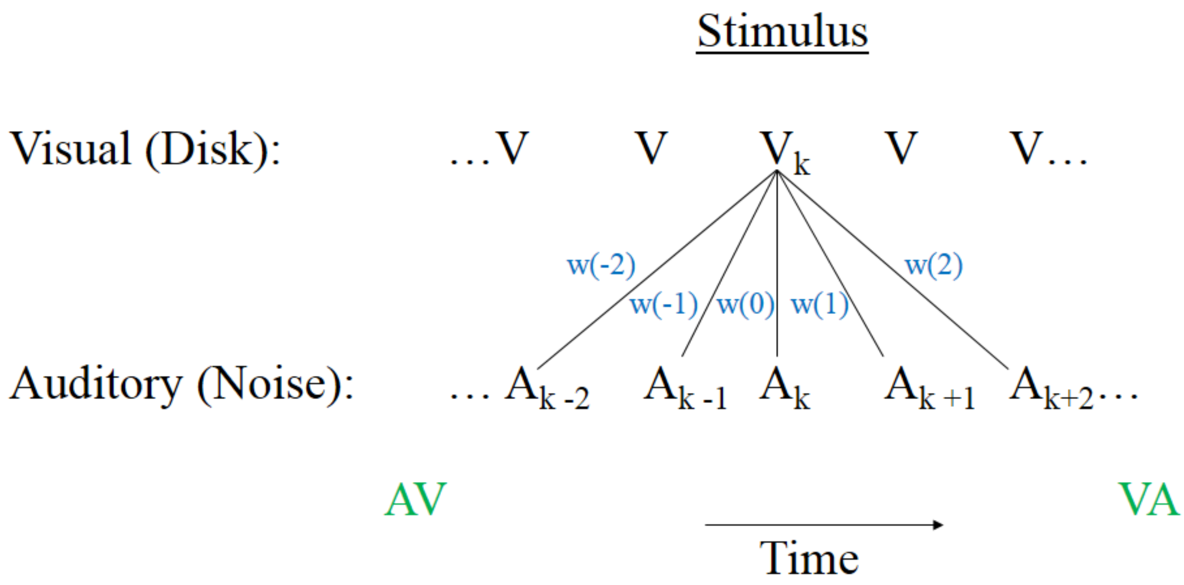


Figure 3.4 *Temporal Binding Parameters*. This is a representation of the visual and auditory components of a stimulus. To explore the temporal binding window, we added 5 parameters to the model in **Equation 3.1**, these parameters are w(-2) through w(2) shown in blue. These parameters reflect the relative strength in which participants weighted the bindings of non-simultaneous disks and noise-bursts. In addition to simultaneous pairings (w(0)), the model accommodates, misbindings in which the auditory component preceded the visual component (AV shown in green), and misbindings in which the visual component preceded the auditory component (VA shown in green). These weights reveal the temporal binding window underlying our participants' decisions.

The difference between a disk and noise-burst in time (separated by tokens) was represented by δ = {-2, -1, 0, 1, 2}. The perceptual influence of the simultaneous event was represented by $w(\delta)$ when $\delta = 0$. If the disk was perceptually paired with the noise-burst that occurred two tokens earlier, this was represented by $w(\delta)$ when $\delta = -2$. If a disk was perceptually paired with the noise-burst that occurred one token earlier, this was represented by $w(\delta)$ when $\delta = -1$. If the disks were being perceptually paired with noise-bursts that occurred later in time, these pairings were represented by $w(\delta)$ when $\delta = 1$ and $\delta = 2$. Each type of pairing was assumed to happen across the majority of the stimulus; the pairings were more complex at the onset and offset of the stimulus stream (see **Equations 3.2 & 3.3**).

In the full model, we have replaced $F(token_k)$ from **Equation 3.1** with $G(k)$ shown below.

$$\sum_{k=1}^{18} G(k) \ * C + Noise < Criterion \tag{3.2}$$

where

$$G(k) = \begin{cases} \sum_{\delta=1-k}^{2} w(\delta) * F\big(V(k), A(k+\delta)\big), & k \leq 2 \\ \sum_{\delta=1-k}^{2} w(\delta) * F\big(V(k), A(k+\delta)\big), & 2 < k < 17 \\ \sum_{\delta=-2}^{18-k} w(\delta) * F(V(k), A(k+\delta)), & k \geq 17 \end{cases} \tag{3.3}$$

where $V(k)$ was the visual component (brightness) of the token at position k in a given stimulus, and $A(k{+}\delta)$ was the auditory component (loudness) of the token at position $k{+}\delta$ of the stimulus. $F(V(k), A(k+\delta))$ was the attention filter (for a given condition) applied to the token-type that was composed of the combination of $V(k)$ and $A(k+\delta)$. **Equation 3.3** demonstrated that all possible visual and auditory bindings up to 2 frames apart were considered in the model. The parameters $w(-2)$,$w(-1)$,..., $w(2)$ for a given participant were fixed across all attention conditions reflecting our assumption that each participant had a fixed temporal binding window that operated in each condition.

Parameter	Constraints	Degrees of Freedom
Attention Filter- $\vec{F}$ (9 parameters per participant per condition)	3. Mean = 0 4. Sum of Squares = 1	7 df (per participant per condition)
Consistency Parameter- C (1 parameter per participant per condition)	none	1 df (per participant per condition)
Criterion Parameter (1 parameter per participant per condition)	none	1 df (per participant per condition)
Temporal Binding Parameters - $w(\delta)$ from $\delta = -2:2$ (5 parameters per participant)	1. Sum of absolute values= 1	4 df (per participant)

Table 3.1 *Full Model Constraints and Degrees of Freedom.* Separate analyses were run for each participant. The full model for each participant had a total of 60 parameters with 49 df[7].

Hierarchical Analysis

In order to evaluate the shared abilities across participants for all conditions, we developed a hierarchical model. This model included a basis of sensitivity functions shared across participants from which all other possible attention filters are derived as linear combinations. We fit this model to all of the data across participants and conditions.

There were 4 basis functions[8], $\overrightarrow{B_1}, \overrightarrow{B_2}, \overrightarrow{B_3},$ and $\overrightarrow{B_4}$, from which all possible attention filters are derived. These basis functions were shared across participants:

$$Basis\ Functions = [\,\overrightarrow{B_1} \quad \overrightarrow{B_2} \quad \overrightarrow{B_3} \quad \overrightarrow{B_4}] \tag{3.6}$$

[7] Except for participant 3 who could not perform the task for the maximum intensity condition. Parameters relevant to that condition were excluded, so the model for that participant had 49 parameters with 40 df.

[8] A model with 5 Basis Functions was also tested, but the 5th Basis Function contributed very little to participant's judgments even though a likelihood-ratio test suggested this model was superior at fitting the data. In addition, the first 4 Basis Functions and α weights remained very stable in both analyses, so we report the 4 Basis Function model.

We assumed that each participant has a fixed strength of each of these functions in their processing system. For each participant, 4 α parameters weighted the strength of each of the $\overrightarrow{B_1}, \overrightarrow{B_2}, \overrightarrow{B_3}$, and $\overrightarrow{B_4}$, basis functions. The participant's αs multiplied by the basis functions gives us a matrix *P,* or Participant-Specific Sensitivity Functions, for each participant:

$$P = [\alpha_1 \overrightarrow{B_1} \quad \alpha_2 \overrightarrow{B_2} \quad \alpha_3 \overrightarrow{B_3} \quad \alpha_4 \overrightarrow{B_4}] \tag{3.7}$$

The model assumed that participants could optimally scale their Participant-Specific Sensitivity Functions, *P*, depending on the selective attention condition. Accordingly, for each participant in each condition, their P matrix was scaled by a set of 4 parameters, $\vec{\beta}$. $\vec{\beta}$ was optimized for each condition, and the elements of $\vec{\beta}$ were not free parameters, they were calculated below.

$$\vec{\beta} = \frac{P^T \vec{T}}{|(P^T \vec{T})|} \tag{3.8}$$

where T was the target filter for a given condition. We assumed that participants could not use their Participant-Specific Sensitivity Functions negatively. Thus, if any optimal β was negative, it was set to zero. Then, the positive the set of βs was forced to have a sum of 1. After the β parameters were applied to the P matrix, we attained the attention filter, $\vec{F}$, for a particular participant in a particular condition.

$$\vec{F} = P * \vec{\beta} \tag{3.9}$$

As in the full model, the attention filter, $\vec{F}$, was applied to each token of the stimulus considering non-simultaneous bindings represented by the Temporal Binding Parameters (see **Equation 3.3**). The standard normal random *Noise* was added, and the total sum was compared to the *Criterion* parameter. The Temporal Binding Parameters in the Hierarchical Model were shared across participants to estimate the shared Temporal Binding Window. The Hierarchical Model excluded the *C,* or consistency, parameters because the α parameters represented the strength of each underlying component or basis function.

Parameter	**Constraints**	**Degrees of Freedom**
Basis Functions $\overrightarrow{B_1}, \overrightarrow{B_2}, \overrightarrow{B_3}$, and $\overrightarrow{B_4}$ (9 parameters in each function = 36 total)	Each of $\overrightarrow{B_1}, \overrightarrow{B_2}, \overrightarrow{B_3}$, and $\overrightarrow{B_4}$ 1. Mean = 0 2. Sum of Squares = 1	28 df
α Parameter (4 parameters per participant = 12 total)	none	12 df
Criterion Parameter (1 parameter per participant per condition = 14 total[9])	none	14 df
Temporal Binding Parameters - $w(\delta)$ from $\delta = -2{:}\,2$ (5 parameters)	Sum of absolute values= 1	4 df

Table 3.2 *Hierarchical Model Constraints and Degrees of Freedom.* The Hierarchical model had a total of 67 parameters with 58 df.

The set of Basis Functions was ordered such that:

- $\overrightarrow{B_1}$ correlated most with the Brightness-only target function
- $\overrightarrow{B_2}$ correlated most with the Loudness-only target function
- $\overrightarrow{B_3}$ correlated most with the Correlatedness target function.

The remaining basis function $\overrightarrow{B_4}$ had no requirements.

[9] This is 14 and not 15 because we did not fit Participant 3's data in the Maximum Intensity Condition. Participant 3 had difficulty performing that particular task.

Model Fitting

The Full Model was fit with a separate Markov Chain Monte-Carlo (MCMC) sampling procedure for each participant. See **Appendix A** for details on the MCMC sampling procedure. The Hierarchical Model fit the data for all participants in one MCMC sampling procedure. All figures display means and 95% credible intervals from the last 80,000 (out of 100,000 collected) samples.

Results

Full Model

Attention Filters. Participants achieved attention filters that closely matched the loudness-only target filter, and their attention filters deviated more from brightness-only target filter (see **Figure 3.5**). In other words, it was more difficult for participants to attend to only brightness and ignore variations in loudness. We observed lower consistency parameter values in the brightness-only condition versus the loudness-only condition. This means that participants adhered to their attention filter more consistently in the loudness-only condition than the brightness-only condition.

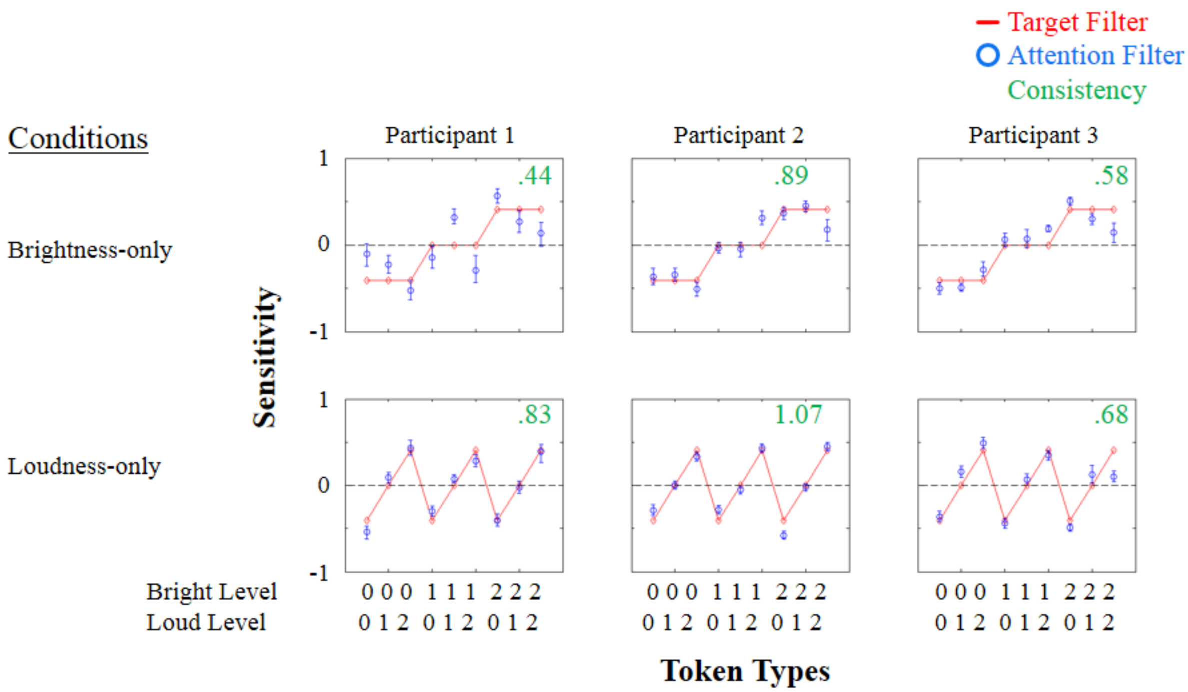


Figure 3.5 *Full Model Results for Single-Modality Attention Conditions*. These are the attention filters (blue) plotted over the target filters (red) for the brightness-only and loudness-only conditions. The attention filter circles are the mean of the posterior and the error bars are 95% credible intervals. The consistency parameter mean estimates are displayed in green. Participants were better at attending to loudness-only versus brightness-only.

In **Figure 3.6**, we plot the attention filters obtained for the conditions that required binding brightness and loudness information. In the correlatatedness condition, the target function assigned equal weight to all three medium brightness tokens; note, however, that for Participants 1 and 3 (and slightly for Participant 2) sensitivity to medium brightness tokens increased with loudness. For the zero and maximum intensity conditions, we see minimal deviations from the target filter. The deviations we see are not surprising or particularly important because participants were able to follow instructions in weighting the particular token-type (either the zero intensity or maximum intensity token-type) highly. Participant 3 had difficulty performing the maximum intensity task, so we did not fit their data for that task.

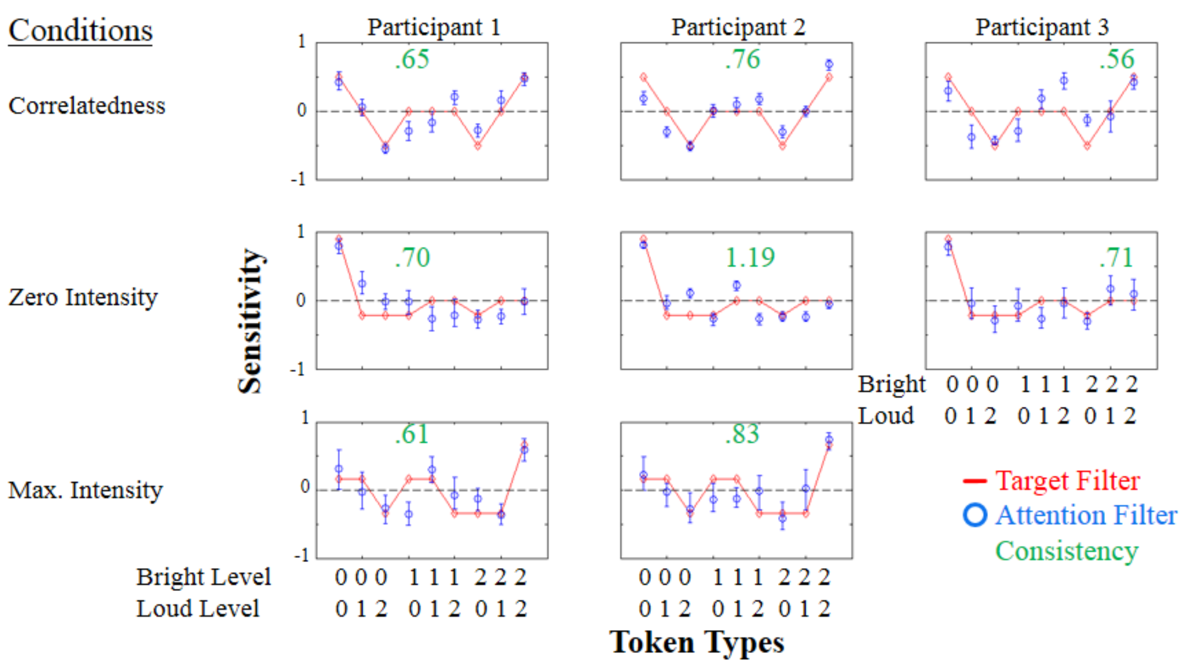


Figure 3.6 *Full Model Results for Binding Attention Conditions*. These are the attention filters (blue) plotted over the target filters (red) for the conditions that required attention to combinations of brightness and loudness information. The attention filter circles are the mean of the posterior and the error bars are 95% credible intervals. The consistency parameter mean estimates are displayed in green. The bottom right plot is missing because participant 3 was unable to perform the maximum intensity task.

Consistency parameters. For all participants, consistency estimates were highest for the Loudness-only and Maximum Intensity conditions. This means participants were most systematic in using their token-weighting filter for these two conditions. Participant 2 had higher average consistency values than Participants 1 and 2.

Participant 1:	Condition	Mean	95% Credible Interval
	Brightness-only	.44	(.39, .49)
	Loudness-only	.83	(.75, .94)
	Correlatedness	.65	(.56, .74)
	Zero Intensity	.70	(.62, .79)
	Maximum Intensity	.61	(.54, .68)
Participant 2:			
	Brightness-only	.89	(.82, .96)
	Loudness-only	1.07	(1.00, 1.14)
	Correlatedness	.76	(.69, .83)
	Zero Intensity	1.19	(1.10, 1.27)
	Maximum Intensity	.83	(.74, .91)
Participant 3:			
	Brightness-only	.58	(.53, .64)
	Loudness-only	.68	(.62, .74)
	Correlatedness	.56	(.48, .65)
	Zero Intensity	.71	(.61, 84)

Table 3.3 *Full Model Consistency Parameter Estimates.* The mean and 95% credible intervals from the last 80,000 MCMC samples.

Temporal Binding Parameters. The estimates for the temporal binding parameters are in **Figure 3.7**. When δ=2 & δ=-2, w(δ) is very close to zero. This represents our temporal binding window. Note that w(1) + w(-1), the sum is approximately equal to w(0). In other words the net influence of loudnesses and brightnesses misbound across a single frame is approximately equal to the net influence of (correctly bound) simultaneous loudnesses and brightnesses.

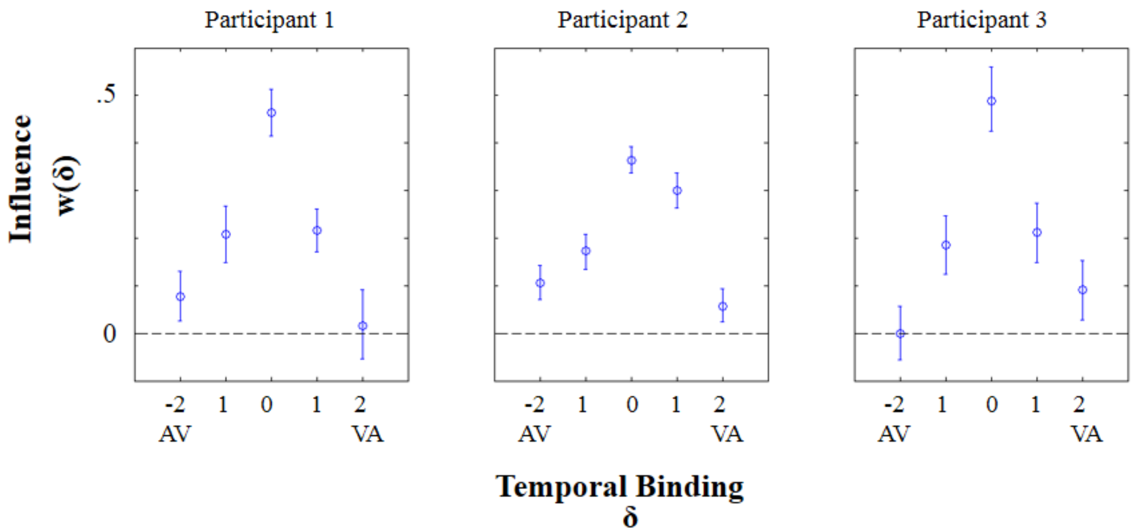


Figure 3.7 *Full Model Temporal Binding Parameters.* For each temporal binding weight, the error bars show the 95% credible intervals, and the circle indicates the mean of the posterior distribution. The different values of δ represent 5 possible bindings across visual and auditory tokens. w(0) represents the weight of the simultaneous disk and noise-burst pairing. Negative values of delta correspond to the perceptual pairings in which the auditory token preceded the visual token (AV), and positive values of delta correspond to the pairings in which the visual token preceded the auditory token (VA).

Each δ corresponds to a time of 83ms (length of each token display), and our temporal binding window is approximately 2 δ's long. This gives us a 167 ms temporal binding window.

Criterion. All criterion parameter estimates were either zero or extremely close to zero (<.5 SD of the standard normal noise).

Hierarchical Model

Basis Functions. The Basis Functions reflected the shared brightness/loudness sensitivity across participants from which all token-weighting filters were derived. The set of Basis Functions was ordered in a way that forced: $\overrightarrow{B_1}$ to be most correlated with the

Brightness-only target filter, $\overrightarrow{B_2}$ to be most correlated with the Loudness-only target filter, and $\overrightarrow{B_3}$ to be most correlated with the Correlatedness target filter.

Figure 3.8 shows the fits for the Basis Functions. $\overrightarrow{B_1}$ showed an interaction with loudness for the high-brightness level tokens; this interaction was also present in the Brightness-only fits in the Full Model. Out of the 4 Basis Functions, $\overrightarrow{B_2}$ most closely matched the target filter (Loudness-only) that it was forced to correlate most with. $\overrightarrow{B_4}$ weighted the zero intensity (blank/silent) token highly and the maximum intensity token lowly.

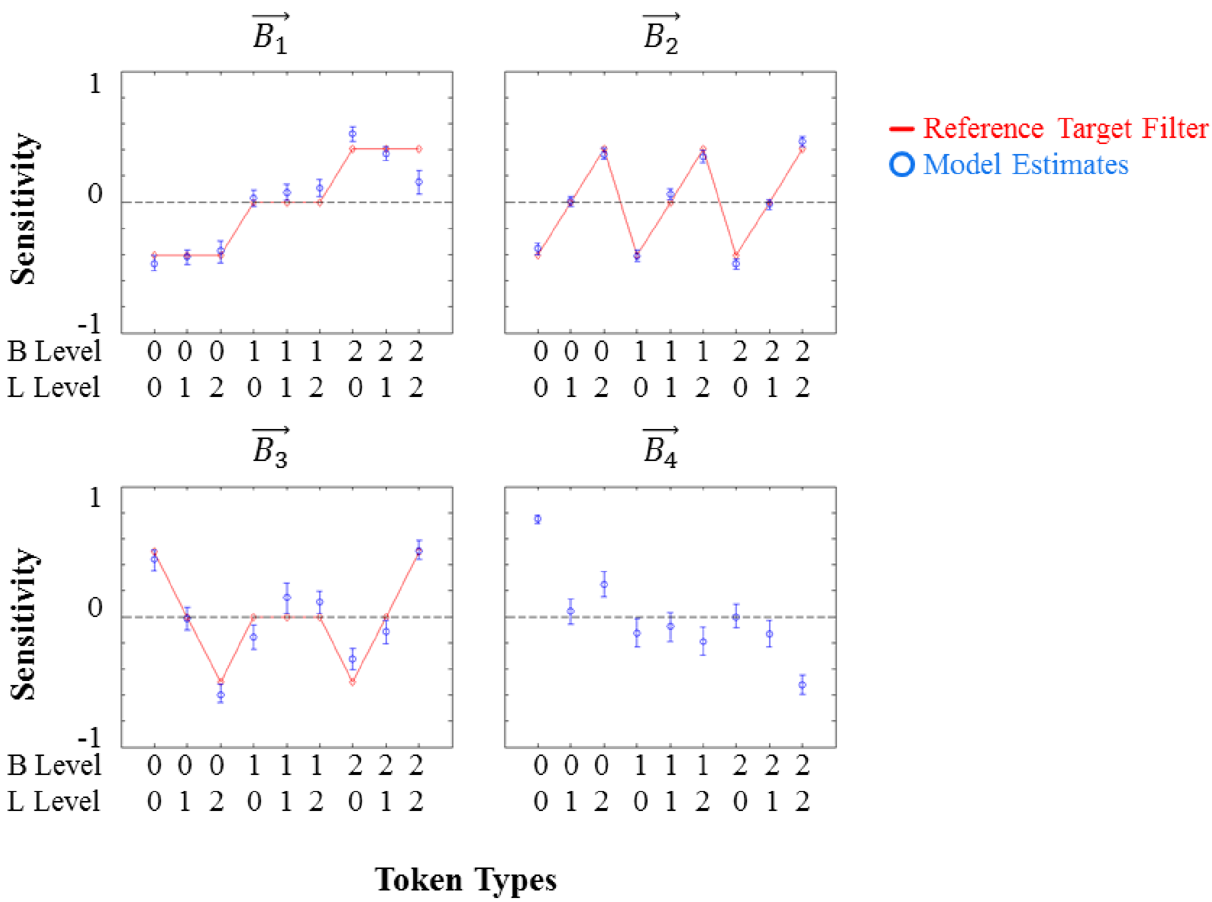


Figure 3.8 *Hierarchical Model Basis Functions.* These Basis functions reflected the shared brightness/loudness sensitivity across participants from which all attention filters were derived. Reference target filters are plotted for the three Basis Functions that were organized based on their correlation with a particular target function.

Alpha parameters. Each participant had 4 α parameters that represented the strength of the representation of each basis function in their processing system. All three participants showed a similar pattern of alpha values in which α_2 and α_4 had the largest influence. Thus, Basis Functions $\overrightarrow{B_2}$ and $\overrightarrow{B_4}$ (Loudness-only and Zero Intensity) were represented most strongly in their processing system (**Figure 3.9**). This pattern aligns with the *Consistency* parameter estimates from the Full Model.

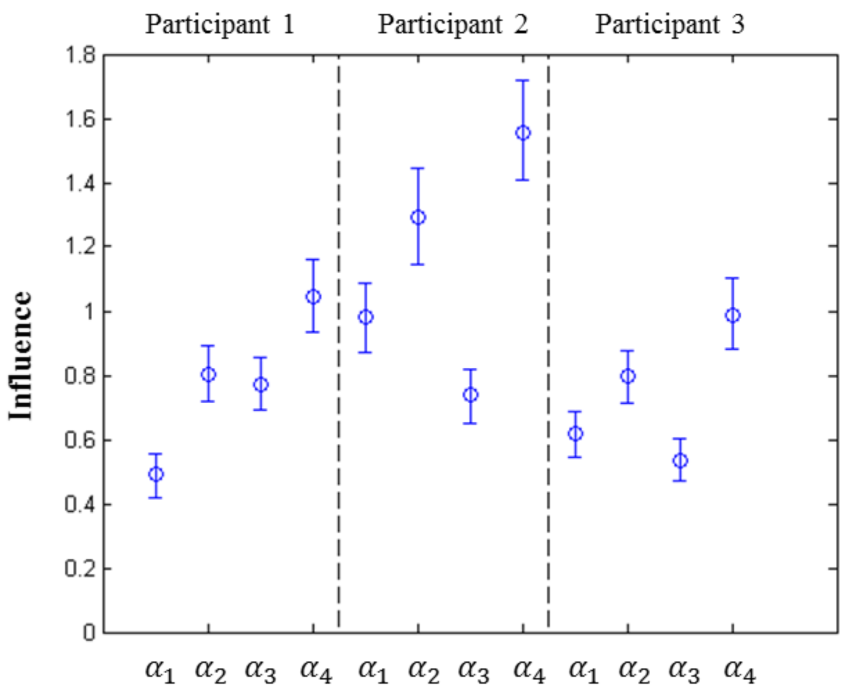


Figure 3.9 *Hierarchical Model Alpha Parameters*. These parameters scaled the influence of the Basis Functions for each participant. The elevated α_2 and α_4 parameters indicate strong representations of $\overrightarrow{\boldsymbol{B_2}}$ and $\overrightarrow{\boldsymbol{B_4}}$ (Loudness-only and Zero-Intensity Functions).

Attention Filters. The attention filters estimated by the Hierarchical Model were plotted over the attention filters from the Full Model (**Figure 3.10**). Each model provided very similar estimates of the attention filters.

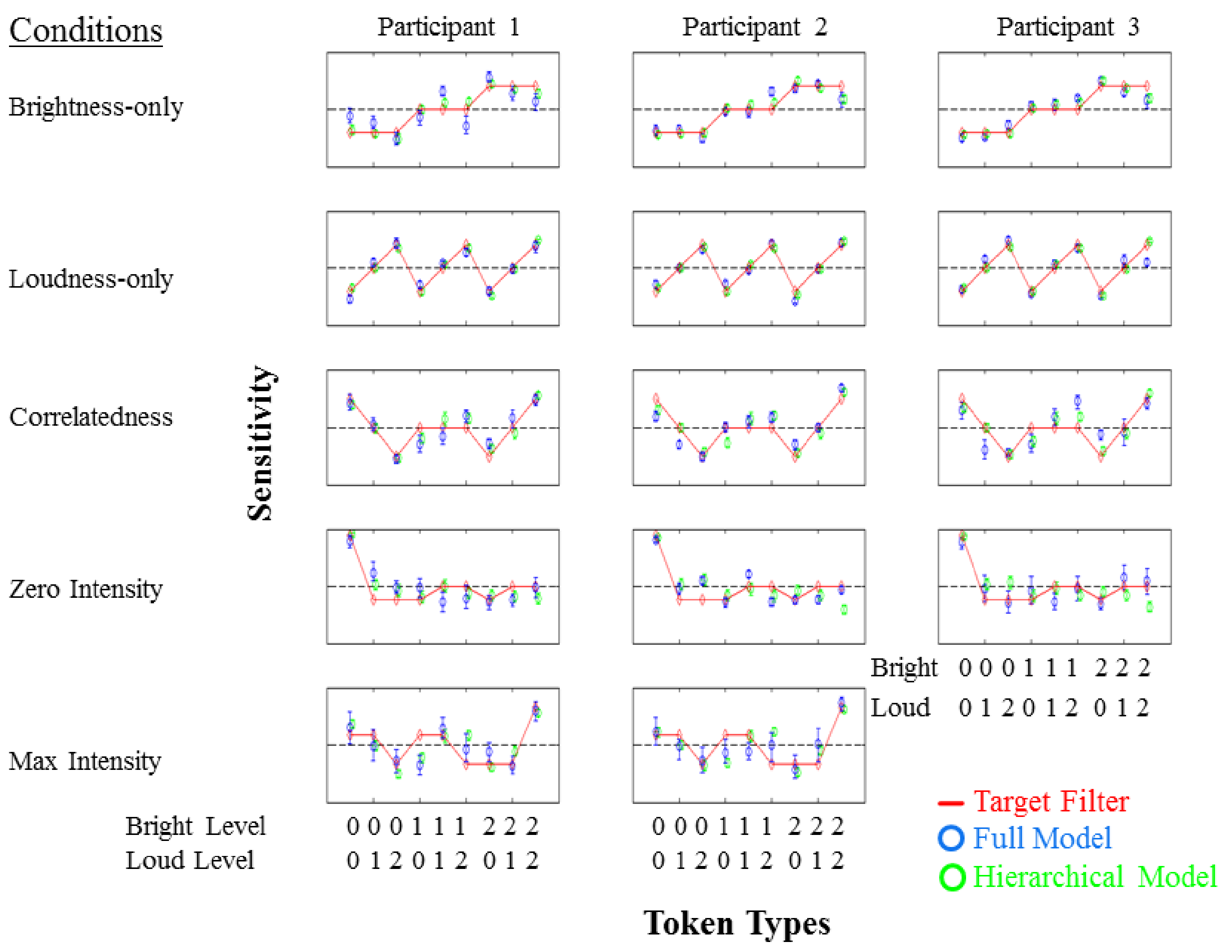


Figure 3.10 *Hierarchical Model Token-Weighting Functions.* The token-weighting functions derived from the Hierarchical model are plotted against with the full model token-weighting functions. There were few deviations between the two model estimates.

Temporal Binding Parameters. These were shared across all participants and conditions in this model. The results (**Figure 3.11**) are very similar to the full model and reflect a confirmation of the temporal limit of binding proposed by Fujisaki and Nishida (2010).

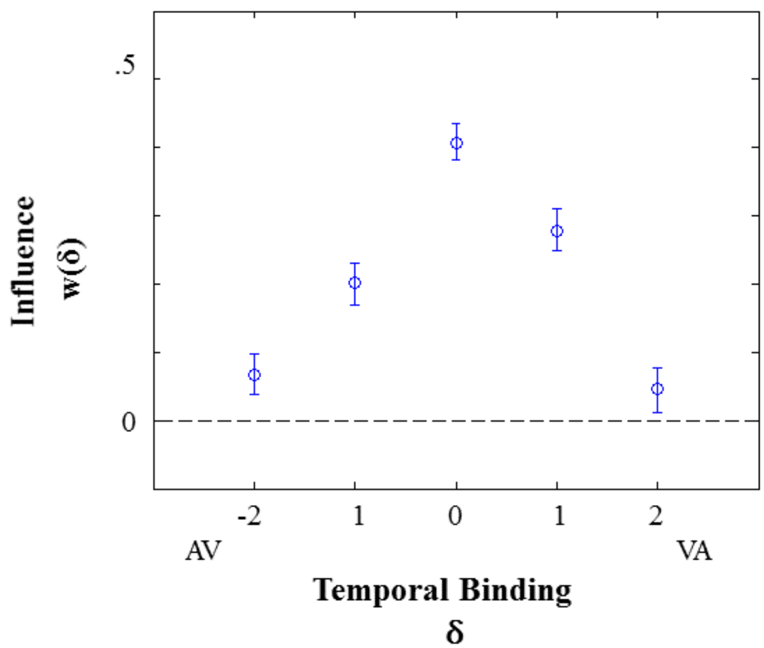


Figure 3.11 *Hierarchical Model Temporal Binding Parameters.* In the Hierarchical Model, this temporal binding window was shared across all participants in all conditions. We find the same result as the full model. This confirms the temporal limit established by Fujisaki & Nishida (2010).

Criterion. As in the full model, all criterion parameter estimates were either zero or extremely close to zero (<.5 SD of the standard normal noise).

Hierarchical Versus Full Model Comparison. The Hierarchical Model accounted for 94.04% of the variance in the trial-by-trial Z-scores predicted by the Full Model.

Discussion

The Limits of Attention on Binding Brightness and Loudness

Participants were able to selectively attend to particular combinations of brightness and loudness- achieve many attention filters. The limits of attention were exposed in each case when the attention filters did not match the target filters. There were constraints on the selective attention process that were not allowing the participant to optimize his/her filter to match the feedback rule.

The full model analysis demonstrated that attending to brightness-only was more difficult than attending to loudness-only. This suggests an asymmetric interaction between loudness and brightness: it was more difficult to decouple variations in brightness from concurrent variations in loudness than it was to decouple variations in loudness from concurrent variations in brightness. Attention to brightness-only was not enough to suppress the multisensory integration between brightness and loudness. It is possible that sound dominates vision when competing for attentional resources. This view is supported by the results of Koelewijn & Theeuwes (2009) who found that auditory attentional capture in a visual stimulus could not be suppressed by top-down attentional control.

In our Hierarchical Analysis, the basis functions shared across participants suggest that there are 4 dimensions of binding brightness & loudness attention. Unsurprisingly, there was a basis function for each brightness-only and loudness only. The third basis function showed attention to correlatedness of brightness and loudness levels. The fourth basis function was a zero intensity (or blank/silent) detector. In other words, we have a mechanism for identifying the absence of brightness and loudness. This final mechanism was represented the strongest for all participants.

Participants were most systematic in the Loudness-only and Zero-Intensity conditions. This was reflected by highest *Consistency* parameter values in the Full Model and the largest α weights in the Hierarchical Model.

Our results contribute to the new exploration of auditory perceptual averaging in a multimodal context. Even in unimodal experiments, auditory perceptual averaging has hardly been explored even though many visual perceptual averaging experiments have been done with size, orientation, location, speed, motion direction, facial identity and emotion (reviewed in Stevenson, 2012). Visual perceptual averaging has also been demonstrated in temporally varying stimuli such as with disc size (Albrecht & Scholl, 2010). Only one recent study by Albrecht (2012) has observed auditory perceptual averaging at all, and it did so in both unimodal and multimodal contexts. In this study, there were temporally varying displays of multiple disc sizes and/or pitches, and participants made magnitude estimations of the average disc size or pitch. The pitch averaging was actually more accurate than disc size, perhaps reflecting the inherent auditory advantage of temporal displays. The present study was the first to investigate loudness averaging in a temporally varying multimodal stimulus, and the robustness of our loudness-only attention filters demonstrate that participants are quite good at loudness averaging.

The Temporal Limit of Binding Brightness and Loudness

Both our Full Model and Hierarchical Model analyses confirmed the Fujisaki & Nishida temporal limit of multimodal binding (2010). This temporal limit translates to a temporal binding window of 167 – 250ms. The present study found a temporal binding window of approximately 167 ms – 332ms; thus, we found converging evidence for this temporal limit of binding. Brightness and loudness are bound with same temporal window for both randomly varying sequences (present study) and oscillating sequences (Fujisaki).

Closing Remarks

We found converging evidence for the previously discovered limit of the multimodal temporal binding window. Participants were able to attend to loudness only and ignore variations in brightness, but they had more trouble attending to brightness only and ignoring loudness. Participants were also able to bind brightness and loudness information in various ways; specifically, our analysis suggests that participants can recruit attentional channels selective for (1) the correlation of brightness and loudness and also for (2) the simultaneous absence of energy in the visual and auditory streams.

REFERENCES

Chapter 1

Bidelman, G. M., Hutka, S., & Moreno, S. (2013). Tone Language Speakers and Musicians Share Enhanced Perceptual and Cognitive Abilities for Musical Pitch: Evidence for Bidirectionality between the Domains of Language and Music. *PLoS ONE*, *8*(4), e60676. doi:10.1371/journal.pone.0060676

Chubb C., Dickson C., Fagan C., Mann D. S., Wright C.E., Silva A. E., Guan H., Gregersen P. K. & Kowalsky E. (2013). Bimodal distribution of performance in discriminating major/minor modes. *Journal of Acoustical Society of America.*

Deutsch, D., Henthorn, T., Marvin, E., & Xu H-S. (2006). Absolute pitch among American and Chinese conservatory students: Prevalence differences, and evidence for a speech-related critical period. *Journal of the Acoustical Society of America*, 119, 719-722

Giuliano, R. J., Pfordresher, P. Q., Stanley, E. M., Narayana, S., & Wicha, N. Y. Y. (2011). Native Experience with a Tone Language Enhances Pitch Discrimination and the Timing of Neural Responses to Pitch Change. *Frontiers in Psychology*, *2*. doi:10.3389/fpsyg.2011.00146

Green, D.M., Swets J.A. (1966) Signal Detection Theory and Psychophysics. New York: Wiley.

Halpern, A. (1984). Perception of structure in novel music. *Memory and Cognition* 12, pp. 163-170.

Halpern, A., Bartlett, J., and Dowling, W. (1998). Perception of mode, rhythm, and contour in unfamiliar melodies: Effects of age and experience. *Music Perception* 15, pp. 335-356.

Jackendoff, R. and Lerdahl, F. (2006). The capacity for music: What is it, and what's special about it? *Cognition* 100, pp. 33–72.

Krumhansl, C. (2002). Music: A link between cognition and emotion. *Current Directions in Psychological Science* 11, pp. 45–50.

Leaver, A. and Halpern, A. (2004). Effects of training and melodic features on mode perception.
Music Perception 22, pp. 117–143.

Lerdahl, F. (2009). Genesis and architecture of the GTTM project. *Music Perception* 26, 42 pp. 187–194.

Miyazaki, K., & Ogawa, Y. (2006). Learning absolute pitch by children. Retrieved from http://www.jstor.org/stable/10.1525/mp.2006.24.1.63

Peretz, I., Champod, S. & Hyde, K. (2003) Varieties of Musical Disorders: The Montreal Battery of Evaluation of Amusia. *Annals of the New York Academy of Sciences*, vol. 999, pp. 58-75.

Russo, F. A., Windell, D. L., & Cuddy, L. L. (2003). Learning the "Special Note": Evidence for a Critical Period for Absolute Pitch Acquisition. *Music Perception*, *21*(1), 119–127. doi:10.1525/mp.2003.21.1.119

Sacks, O. W. (2007). *Musicophilia: Tales of music and the brain*. New York: Alfred A. Knopf.

Chapter 2

Barnes, R., & Jones, M. R. (2000). Expectancy, Attention, and Time. *Cognitive Psychology*, *41*(3), 254–311. doi:10.1006/cogp.2000.0738

Bigand, E. (1997). Perceiving musical stability: The effect of tonal structure, rhythm, and musical expertise. *Journal of Experimental Psychology: Human Perception and Performance*, *23*(3), 808.

Boltz, M. (1989). Rhythm and "good endings": Effects of temporal structure on tonality judgments. *Perception & Psychophysics*, *46*(1), 9–17.

Crane F. (1976). Review of Diabolus in musica: Studien zur Ikonographie der Musik im Mittelalter by Reinhold Hammerstein. *Journal of the American Musicological Society, 29* (2), 300-303.

Chubb C., Dickson C., Fagan C., Mann D. S., Wright C.E., Silva A. E., Guan H., Gregersen P. K. & Kowalsky E. (2013). Bimodal distribution of performance in discriminating major/minor modes. *Journal of Acoustical Society of America.*

Dawe, L. A., Platf, J. R., & Racine, R. J. (1993). Harmonic accents in inference of metrical structure and perception of rhythm patterns. *Perception & Psychophysics*, *54*(6), 794–807.

Dawe, L. A., Platt, J. R., & Racine, R. J. (1995). Rhythm perception and differences in accent weights for musicians and nonmusicians. *Perception & Psychophysics*, *57*(6), 905–914.

Drake, C., Jones, M. R., & Baruch, C. (2000). The development of rhythmic attending in auditory sequences: attunement, referent period, focal attending. *Cognition*, *77*(3), 251–288.

Fraisse, P. (1982). Rhythm and tempo. In D. Deutsch (Ed.), *The psychology of music* (pp. 149-180). New York: Academic Press.

Halpern, A. R., Bartlett, J. C., & Dowling, W. J. (1998). Perception of Mode, Rhythm, and Contour in Unfamiliar Melodies: Effects of Age and Experience. *Music Perception: An Interdisciplinary Journal*, *15*(4), 335–355. doi:10.2307/40300862

Halpern, A. R., Martin, J. S., & Reed, T. D. (2008). An ERP Study of Major-Minor Classification in Melodies. *Music Perception: An Interdisciplinary Journal*, *25*(3), 181–191. doi:10.1525/mp.2008.25.3.181

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

Jones, M. R., Boltz, M., & Kidd, G. (1982). Controlled attending as a function of melodic and temporal context. *Perception & Psychophysics*, *32*(3), 211–218.

Jones, M. R., Johnston, H. M., & Puente, J. (2006). Effects of auditory pattern structure on anticipatory and reactive attending. *Cognitive Psychology*, *53*(1), 59–96. doi:10.1016/j.cogpsych.2006.01.003

Jones, M. R., Moynihan, H., MacKenzie, N., & Puente, J. (2002). Temporal Aspects of Stimulus-Driven Attending in Dynamic Arrays. *Psychological Science, 13*(4): 313-319.

Jones, M. R., & Ralston, J. T. (1991). Some influences of accent structure on melody recognition. *Memory & Cognition*, *19*(1), 8–20.

Jones, M. R., Summerell, L., & Marshburn, E. (1987). Recognizing melodies: A dynamic interpretation. *The Quarterly Journal of Experimental Psychology Section A*, *39*(1), 89–121. doi:10.1080/02724988743000051

Kidd, G., Boltz, M., & Jones, M. R. (1984). Some Effects of Rhythmic Context on Melody Recognition. *The American Journal of Psychology*, *97*(2), 153. doi:10.2307/1422592

Krumhansl CL. (2000) Rhythm and pitch in music cognition. *Psychological Bulletin 126*(1):159-79.

Krumhansl, C. L., & Cuddy, L. L. (2010). A theory of tonal hierarchies in music. In M. R. Jones, R. R. Fay, & A. N. Popper (Eds.), Music perception. New York: Springer.

Lerdahl, F., & Jackendoff, R. S. (1983). *A Generative Theory of Tonal Music*. MIT Press.

Fux, J. J. (1965). The Study of Counterpoint from Johann Joseph Fux's Gradus Ad Parnassum. W. W. Norton & Company.

Palmer, C., & Krumhansl, C. L. (1987a). Independent temporal and pitch structures in determination of musical phrases. *Journal of Experimental Psychology: Human Perception and Performance*, *13*(1), 116.

Palmer, C., & Krumhansl, C. L. (1987b). Pitch and temporal contributions to musical phrase perception: Effects of harmony, performance timing, and familiarity. *Perception & Psychophysics*, *41*(6), 505–518.

Peretz, I., & Kolinsky, R. (1993). Boundaries of separability between melody and rhythm in music discrimination: A neuropsychological perspective. *The Quarterly Journal of Experimental Psychology Section A*, *46*(2), 301–325. doi:10.1080/14640749308401048

Schellenberg, E. G., Krysciak, A. M., & Campbell, R. J. (2000). Perceiving Emotion in Melody: Interactive Effects of Pitch and Rhythm. *Music Perception: An Interdisciplinary Journal*, *18*(2), 155–171. doi:10.2307/40285907

Chapter 3

Albrecht, A. R., & Scholl, B. J. (2010). Perceptually Averaging in a Continuous Visual World: Extracting Statistical Summary Representations Over Time. *Psychological Science*, *21*(4), 560–567. doi:10.1177/0956797610363543

Albrecht, Alice R., Scholl, B. J., & Chun, M. M. (2012). Perceptual averaging by eye and ear: Computing summary statistics from multimodal stimuli. *Attention, Perception, & Psychophysics*, *74*(5), 810–815. doi:10.3758/s13414-012-0293-0

Arieh, Y., & Marks, L. E. (2008). Cross-modal interaction between vision and hearing: A speed-accuracy analysis. *Perception & Psychophysics*, *70*(3), 412–421. doi:10.3758/PP.70.3.412

Ben-Artzi E., Marks L. E. (1995). Visual-auditory interaction in speeded classification: role of stimulus difference. *Perception & Psychophysics 57*(8):1151-62.

Colonius, H., & Diederich, A. (2004). Multisensory interaction in saccadic reaction time: a time-window-of-integration model. *Journal of Cognitive Neuroscience*, *16*(6), 1000–1009.

De Fockert, J., & Wolfenstein, C. (2009). Rapid extraction of mean identity from sets of faces. *The Quarterly Journal of Experimental Psychology*, *62*(9), 1716–1722. doi:10.1080/17470210902811249

Fujisaki, W., & Nishida, S. (2010). A common perceptual temporal limit of binding synchronous inputs across different sensory attributes and modalities. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1692), 2281–2290. doi:10.1098/rspb.2010.0243

Koelewijn, T., Bronkhorst, A., & Theeuwes, J. (2009). Auditory and visual capture during focused visual attention. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(5), 1303–1315. doi:10.1037/a0013901

Lippert, M., Logothetis, N. K., & Kayser, C. (2007). Improvement of visual contrast detection by a simultaneous sound. *Brain Research*, *1173*, 102–109. doi:10.1016/j.brainres.2007.07.050

Marks, L. E. (1989). On cross-modal similarity: The perceptual structure of pitch, loudness, and brightness. *Journal of Experimental Psychology: Human Perception and Performance*, *15*(3), 586.

Marks, L. E., Ben-Artzi, E., & Lakatos, S. (2003). Cross-modal interactions in auditory and visual discrimination. *International Journal of Psychophysiology*, *50*(1), 125–145.

Odgaard, E. C., Arieh, Y., & Marks, L. E. (2003). Cross-modal enhancement of perceived brightness: sensory interaction versus response bias. *Perception & Psychophysics*, *65*(1), 123–132.

Shams L., Kamitani Y., Shimojo S. (2000). Illusions: What you see is what you hear. *Nature*. 408(6814):788.

Shams, L., & Kim, R. (2010). Crossmodal influences on visual perception. *Physics of Life Reviews*, *7*(3), 269–284. doi:10.1016/j.plrev.2010.04.006

Stein, B. E., London, N., Wilkinson, L. K., & Price, D. D. (1996). Enhancement of perceived visual intensity by auditory stimuli: a psychophysical analysis. *Journal of Cognitive Neuroscience*, *8*(6), 497–506.

Stevenson, R. A., Zemtsov, R. K., & Wallace, M. T. (2012). Individual differences in the multisensory temporal binding window predict susceptibility to audiovisual illusions. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(6), 1517–1529. doi:10.1037/a0027339

Talsma, D., Doty, T. J., & Woldorff, M. G. (2006). Selective Attention and Audiovisual Integration: Is Attending to Both Modalities a Prerequisite for Early Integration? *Cerebral Cortex*, *17*(3), 679–690. doi:10.1093/cercor/bhk016

Talsma, D., Senkowski, D., Soto-Faraco, S., & Woldorff, M. G. (2010). The multifaceted interplay between attention and multisensory integration. *Trends in Cognitive Sciences*, *14*(9), 400–410. doi:10.1016/j.tics.2010.06.008

Vroomen, J., Bertelson, P., & De Gelder, B. (2001). The ventriloquist effect does not depend on the direction of automatic visual attention. *Perception & Psychophysics*, *63*(4), 651–659.

APPENDIX A: Chapter 2 Stimuli: Note-Count Vectors

These are the possible note-count vectors for the different difficulties in the experiment of Chapter 2.

Note-count vectors when Difficulty = 1

G_5	G♯	A	A♯	B	C	C♯	D	D♯	E	F	F♯	G_6
3	0	0	0	4	0	0	2	0	4	0	0	2
3	0	0	0	3	0	0	2	1	4	0	0	2
3	0	0	0	4	0	0	2	1	3	0	0	2
3	0	0	1	3	0	0	2	0	4	0	0	2
3	0	0	1	4	0	0	2	0	3	0	0	2
3	0	0	1	3	0	0	2	1	3	0	0	2
2	0	0	0	4	0	0	2	0	4	0	0	3
2	0	0	0	3	0	0	2	1	4	0	0	3
2	0	0	0	4	0	0	2	1	3	0	0	3
2	0	0	1	3	0	0	2	0	4	0	0	3
2	0	0	1	4	0	0	2	0	3	0	0	3
2	0	0	1	3	0	0	2	1	3	0	0	3
3	0	0	4	0	0	0	2	4	0	0	0	2
3	0	0	4	1	0	0	2	3	0	0	0	2
3	0	0	4	0	0	0	2	3	1	0	0	2
3	0	0	3	1	0	0	2	4	0	0	0	2
3	0	0	3	0	0	0	2	4	1	0	0	2
3	0	0	3	1	0	0	2	3	1	0	0	2
2	0	0	4	0	0	0	2	4	0	0	0	3
2	0	0	4	1	0	0	2	3	0	0	0	3
2	0	0	4	0	0	0	2	3	1	0	0	3
2	0	0	3	1	0	0	2	4	0	0	0	3
2	0	0	3	0	0	0	2	4	1	0	0	3
2	0	0	3	1	0	0	2	3	1	0	0	3

Note-count vectors when Difficulty = 2

G_5	G♯	A	A♯	B	C	C♯	D	D♯	E	F	F♯	G_6
2	2	0	0	2	0	0	2	0	2	2	2	1
3	1	0	0	2	0	0	2	0	2	2	2	1
3	0	1	0	2	0	0	2	0	2	2	2	1
3	0	0	0	2	0	0	2	0	2	1	2	3
3	0	0	0	2	0	0	2	0	2	2	1	3
3	0	0	0	2	0	0	2	0	2	2	2	2

2	0	2	2	0	2	2	0	2	0	0	0	3
1	1	2	2	0	2	2	0	2	0	0	0	3
1	2	1	2	0	2	2	0	2	0	0	0	3
1	2	2	2	0	2	2	0	2	0	1	0	1
1	2	2	2	0	2	2	0	2	0	0	1	1
1	2	2	2	0	2	2	0	2	0	0	0	2

Note-count vectors when Difficulty = 3

G_5	**G♯**	**A**	**A♯**	**B**	**C**	**C♯**	**D**	**D♯**	**E**	**F**	**F♯**	**G_6**
2	2	0	0	2	0	0	2	0	2	2	2	1
3	1	0	0	2	0	0	2	0	2	2	2	1
3	0	1	0	2	0	0	2	0	2	2	2	1
3	0	0	1	2	0	0	2	0	2	2	2	1
3	0	2	1	2	2	0	0	0	2	0	0	3
3	2	0	0	1	0	2	0	0	2	0	2	3
3	2	0	0	2	0	2	2	1	2	0	0	1
3	0	2	0	2	2	2	0	0	1	0	2	1
3	0	0	0	2	0	0	2	0	1	2	2	3
3	0	0	0	2	0	0	2	0	2	1	2	3
3	0	0	0	2	0	0	2	0	2	2	1	3
3	0	0	0	2	0	0	2	0	2	2	2	2
2	0	2	2	0	2	2	0	2	0	0	0	3
1	1	2	2	0	2	2	0	2	0	0	0	3
1	2	1	2	0	2	2	0	2	0	0	0	3
1	2	2	1	0	2	2	0	2	0	0	0	3
1	2	0	1	0	0	2	2	2	0	2	2	1
1	0	2	2	1	2	0	2	2	0	2	0	1
1	0	2	2	0	2	0	0	1	0	2	2	3
1	2	0	2	0	0	0	2	2	1	2	0	3
1	2	2	2	0	2	2	0	2	1	0	0	1
1	2	2	2	0	2	2	0	2	0	1	0	1
1	2	2	2	0	2	2	0	2	0	0	1	1
1	2	2	2	0	2	2	0	2	0	0	0	2

Note-count vectors when Difficulty = 4

G_5	**G♯**	**A**	**A♯**	**B**	**C**	**C♯**	**D**	**D♯**	**E**	**F**	**F♯**	**G_6**
3	0	2	0	2	2	2	0	0	0	1	2	1
3	0	0	0	2	0	0	2	0	2	1	2	3
3	2	2	0	0	0	2	0	0	2	1	0	3
3	2	0	2	2	2	0	0	0	2	1	0	1
3	0	2	0	2	2	2	0	0	0	2	1	1
3	0	0	0	2	0	0	2	0	2	2	1	3

3	2	2	0	0	0	2	0	0	2	0	1	3
3	2	0	2	2	2	0	0	0	2	0	1	1
3	0	2	0	2	2	2	0	0	0	2	0	2
3	0	0	0	2	0	0	2	0	2	2	2	2
3	2	2	0	0	0	2	0	0	2	0	2	2
3	2	0	2	2	2	0	0	0	2	0	0	2
2	0	2	2	0	2	2	0	2	0	0	0	3
2	0	0	2	2	2	0	2	2	0	2	0	1
2	0	2	0	0	0	2	2	2	0	2	2	1
2	0	0	2	0	0	0	2	2	2	0	2	3
1	1	2	2	0	2	2	0	2	0	0	0	3
1	1	0	2	2	2	0	2	2	0	2	0	1
1	1	2	0	0	0	2	2	2	0	2	2	1
1	1	0	2	0	0	0	2	2	2	0	2	3
1	2	1	2	0	2	2	0	2	0	0	0	3
1	0	1	2	2	2	0	2	2	0	2	0	1
1	2	1	0	0	0	2	2	2	0	2	2	1
1	0	1	2	0	0	0	2	2	2	0	2	3

APPENDIX B: Markov Chain Monte Carlo Sampling Procedure

For the Bayesian modeling procedures in CH 2 & 3, a Markov Chain Monte Carlo (MCMC) simulation was used to estimate the joint posterior density of the parameter-vector used to fit the data. This MCMC estimation method involved picking a random point in the multi-dimensional parameter space (with large uninformative priors), and computing the log-likelihood of that point. Then a Gaussian jump was made to the next *candidate* sample and the log-likelihood was computed of that candidate. If the *candidate* log-likelihood was larger than the *previous* candidate sample, then we added the new candidate to the list with probability of the likelihood ratio of *candidate/previous*. Otherwise, a copy of the previous element was added to the posterior. After each 1,000 samples, a singular value decomposition adjusted the parameter vector to an appropriate direction in the multidimensional parameter space. 100,000 samples were collected, and the last 80,000 samples reflected a stable estimate of the posterior density (unless otherwise noted in each analysis). All the parameter means and 95% credible intervals were pulled from the last 80,000 samples. See Appendix 4 in "A method for analyzing the dimensions of preattentive visual sensitivity" by Chubb, Scofield, Chaio & Sperling (2012)[10] for more details on this analysis technique.

[10] Chubb, C., Scofield, I., Chiao, C.-C., & Sperling, G. (2012). A method for analyzing the dimensions of preattentive visual sensitivity. *Journal of Mathematical Psychology*, *56*(6), 427–443.