UNIVERSITY OF CALIFORNIA,
IRVINE

# Trying Not to Fall Flat: Cognitive and Neural Investigations of Major/Minor Musical Processing

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Cognitive Sciences

by

Solena Davine Mednicoff

Dissertation Committee:
Professor Charles Chubb, Chair
Professor Emily Grossman
Professor Gregory Hickok

2021

# DEDICATION

This dissertation is dedicated to my parents, my friends and family, and importantly, myself. We did this together.

To my mother, thank you for instilling your love for music in me to help bring this dissertation to fruition. To my father, thank you for instilling your passion for education, and teaching me the value in learning.

For all of the endless and doubtful nights, this dissertation is a true testament and dedication to myself, and those I will continue to lift up in the future.

We persisted.

# TABLE OF CONTENTS

# LIST OF FIGURES

Page

# LIST OF TABLES

# ACKNOWLEDGMENTS

I would like to thank my dissertation committee for their contributions to this dissertation and my career trajectory. Without your invaluable input, this dissertation would not be possible.

Thank you to my advisor, Professor Charles Chubb, for the years you've dedicated to help create a space for me to learn and grow as an independent researcher. I would also like to thank Professor Gregory Hickok for his support and encouragement.

I am also deeply grateful for Professors Virginia Richards and Emily Grossman for their years of dedication, helpful advice, conversations, and wisdom. Most importantly, thank you for being my role models. I would not have grown as much as I have as an independent researcher without your guidance.

My deepest gratitude also goes to each undergraduate student who I've mentored and who has helped contribute to this dissertation. Your time and energy made this work fun, enjoyable, and worthwhile.

I would also like to acknowledge and thank the UCI Center for Hearing Research for their training and funding support for this dissertation under the NIH Training Grant T32-DC010775.

Lastly, I would like to acknowledge the Journal of the Acoustical Society of America for their permission to incorporate my work published in their journal for my dissertation.

# CURRICULUM VITAE

## Solena Davine Mednicoff

### EDUCATION

**Doctor of Philosophy in Cognitive Sciences** **2021**
with an emphasis in Cognitive Neuroscience *Irvine, CA*
University of California, Irvine

**Master of Science in Cognitive Neurocience** **2017**
University of California, Irvine *Irvine, CA*

**Bachelor of Science in Neuroscience** **2015**
University of Nevada, Reno *Reno, NV*

### RESEARCH EXPERIENCE

**Graduate Student Researcher** **2019–2021**
UCI Graduate Division *Irvine, California*

**Graduate Student Researcher** **2015–2021**
UCI Cognitive Sciences Department *Irvine, California*

**Center for Hearing Research Predoctoral Fellow** **2016–2018**
UCI Center for Hearing Research *Irvine, California*

### TEACHING EXPERIENCE

**Instructor of Record** **2018–2018**
University of California Irvine *Irvine, CA*

Courses: Psychology Fundamentals 9B/ Psychology and Social Behaviors 11B

**Teaching Assistant** **2015–2019**
University of California, Irvine *Irvine, CA*

Courses (# of quarters taught): Abnormal Psychology (1),
Writing about Memory (1), Psychology of Social Behaviors (1),
Psychology Research Methods (1), Psychology and Social Behaviors 11A (2)

## JOURNAL PUBLICATIONS

**Many listeners cannot discriminate major or minor tone-scrambles regardless of presentation rate.** **2018**
The Journal of the Acoustical Society of America.

Authors: **Mednicoff, S.**, Mejia, S., Rashid, J., and Chubb, C.

**Can training improve performance in classifying major versus minor tone-scrambles?** **Submitted 2021**
The Journal of the Acoustical Society of America.

Authors: **Mednicoff, S.** and Chubb, C.

## CONFERENCE PUBLICATIONS

**The effects of tone duration in discriminating major/minor modes.** **July 2018**
The 15th International Conference for Music Perception and Cognition

Authors: **Mednicoff, S.**, Mejia, S., Rashid, J., and Chubb, C.

## TALKS

Chubb, C., Dean, T., **Mednicoff, S.**, Ho, J., Waz, S., Douthitt C., Comishen, K., Adler, S. (2019, August) What tone-scramble experiments reveal. Talk presented by C. Chubb at the 2019 Society for Music Perception and Cognition Conference, New York City, NY.

**Mednicoff, S.** Mejia, S., Rashid, J., and Chubb, C. (2018, July). The effects of tone duration in discriminating major/minor modes. Talk presented at the 15th International Conference for Music Perception and Cognition, Montreal, Quebec, CA.

**Mednicoff, S.** (2018, January). The effects of tone duration in discriminating major/minor modes. Third Year Pre-advancement talk at the Cognitive Sciences Department, UC Irvine.

**Mednicoff, S.** (2017, January). Bimodal Distribution of Performance in Discriminating Major/Minor Modes. Journal Club Presentation at Center for Hearing Research, UC Irvine.

**Mednicoff, S.** (2016, April). Your Brain as a Musician. Speech presented at the Las Vegas Academy Piano Institute 2016 Sharing is Caring Symposium, Las Vegas, NV.

## POSTERS

**Mednicoff, S.** and Chubb, C. (2019, Sept). *Discriminating Major/Minor Modes Revisited: Does Training Help?* Poster presented at the SoCal Hearing Conference, Irvine, CA.

Huynh, M., Hwang, S., Ho, J., **Mednicoff, S.**, Chubb, C., Hickok, G. (2019, May). *Assess-*

*ing Emotional Processing in People with Autism Spectrum Disorder.* Poster presented at the Undergraduate Research Symposium, Irvine, CA.

Pham, N., Lopez, E., Gonzalez, K., Ho, J., **Mednicoff, S.**, and Chubb, C. (2018, May). *Pitch Discrimination in Tone Sequences and Speech Production.* Poster presented at the Undergraduate Research Symposium, Irvine, CA.

**Mednicoff, S.** and Chubb, C. (2017, Aug). *Discriminating Major/Minor Modes Revisited: Does Training Help?* Poster presented at the Society for Music Perception and Cognition conference, San Diego, CA.

Mejia, S., **Mednicoff, S.**, Rashid, J., and Chubb, C. (2017, May). *Effects of Tone Duration in Tonality Perception.* Poster presented at the Undergraduate Research Symposium, Irvine, CA.

**Mednicoff, S.**, Skiba, R.M., and Webster, M. (2015, Aug). *Ensemble coding within and across musical instruments.* Poster presented at the 2015 Society for Music Cognition and Perception Conference, Nashville, TN.

Ara, J., **Mednicoff, S.**, and Webster, M. (2015, May). *Adaptation and compensation in anomalous trichromacy.* Poster presented at the 2015 Visual Sciences Society Conference, St. Pete Beach, FL.

**Mednicoff, S.**, Skiba, R.M., and Webster, M. (2015, April). *Ensemble coding in music perception.* Poster presented at the 2015 Undergraduate Research Poster Conference, Reno, NV.

**Mednicoff, S.**, Skiba, R.M., and Webster, M. (2015, April). *Ensemble coding within and across musical instruments.* Poster presented at the 2015 Western Regional Honors Conference, Reno, NV. (1st Place in Poster Competition)

Banks, S., **Mednicoff, S.** and Bernick, C. (2015, Feb). *Verbal Fluency and Brain Health in Boxers and Mixed Martial Arts Fighters.* Poster presented at the 2015 International Neuropsychological Society Annual Meeting Denver, CO.

Sternberg, H., **Mednicoff, S.**, Goto, K., Mac, A., Tregillus, K., Kingdom F.A.A., and Webster, M.A. (2014, May). *Painted faces: Misperceiving shading as pigmentation.* Poster presented at the Psi Chi, Undergraduate Research Fair, Reno. NV. (1st Place in Poster Competition)

## AWARDS & ACHIEVEMENTS

**Graduate Student Research Fellowship** **2019 - 2021**
UC Irvine, Graduate Division
**Associate Dean Fellowship** **Fall 2018**
UC Irvine, School of Social Sciences
**Graduate Student Travel Award** **May 2018**

UC Irvine, School of Social Sciences
**SEMPRE Student Travel Award** **April 2018**
International Conference for Music Perception & Cognition
**Center for Hearing Research Project Funding** **October 2017**
UC Irvine, Center for Hearing Research
**NIH Sponsored fMRI Training Program** **August 2017**
University of Michigan, Ann Arbor
**Rapid Grant Award** **Fall 2016**
UC Music Experience Research Community Initiative
**T32 NIH-NRSA Predoctoral Training Fellowship** **2016 - 2018**
UC Irvine, Center for Hearing Research
**1st Place Poster Award 2015 Winner** **April 2015**
Western Regional Honors Conference
**Honors Undergraduate Research Award** **October 2014**
University of Nevada, Reno
**First Place Poster Psi Chi Research Fair** **May 2014**
University of Nevada, Reno
**Millenium Scholarship** **2011 - 2015**
University of Nevada, Reno
**Barringere Scholarship for music performance** **2011 - 2015**
University of Nevada, Reno
**Albert, Henry, & Edith Scholarship** **2011 - 2013**
University of Nevada, Reno

## RESEARCH SUPPORT

**Center for Hearing Research Project Funding:** Neural Substrates of Musical Sensitivity: Major and Minor Discrimination as a Function of Musical Experience 10/31/2017 - 06/01/2021
Total costs: $10,000
Role: Co-Investigator as a graduate student (PI: Charles Chubb)

**UC Music Experience Research Community Initiative (UC MERCI) Pilot Grant:** *The Rhythm of Perception: Walking Entrainment and Tonality's Relationship to Speech and Language* 10/01/2016 - 03/31/2017
Total costs: $5,000
Role: Co-Investigator as a graduate student (PI: Charles Chubb and Greg Hickok)

## SKILLS

| | |
|---|---|
| **Programming Languages** | Matlab, LaTeX, Praat, R, Python |
| **Software & Tools** | BrainVoyager, Microsoft Office, SPSS, JASP, Audacity, Photoshop |
| **Languages** | English (native), Spanish (conversational) |

## PROFESSIONAL AND DEPARTMENTAL SERVICE

**Member, Communications Committee**
UCI Center for Neurobiology of Learning and Memory Brain Ambassadors *2018 - 2021*
**Guest Panelist**
Psychology Fundamentals 9B *2019*
**Chair of Session**
15th International Conference on Music Perception and Cognition *2018*
**Guest Editor**
Center for Hearing Research Sound Minds Newsletter *2017*
**Vice President**
Nu Rho Psi, University of Nevada, Reno *2014 - 2015*

## PROFESSIONAL MEMBERSHIP

Society for Music Perception and Cognition
Society for Education and Music Psychology Research
International Society for Music Perception and Cognition
International Society for Music Information Retrieval
Women in Music Information Retrieval (WiMIR)
Nu Rho Psi - International Honors Society for Neuroscience

# ABSTRACT OF THE DISSERTATION

Trying Not to Fall Flat: Cognitive and Neural Investigations of Major/Minor Musical Processing

By

Solena Davine Mednicoff

Doctor of Philosophy in Cognitive Sciences

University of California, Irvine, 2021

Professor Charles Chubb, Chair

Musical emotions have become a facet in music perception that is ubiquitously understood, yet difficult to define. Music holds the unique ability to communicate many emotions as well as change one's emotional state. For example, major and minor modes hold a special connection in music that drives emotional meaning. Research shows that many listeners hear major and minor music as sounding "happy" and "sad," respectively; however, other research shows that many listeners (70%) cannot discriminate between the major and minor musical modes (Chubb et al., 2013; Dean and Chubb, 2017). Thus, this dissertation uses behavioral and functional magnetic resonance imaging (fMRI) experiments to investigate the difference between those low-performing listeners who cannot and those high-performing listeners who can discriminate between major and minor musical stimuli. Studies that have demonstrated this divide in performance have used stimuli called tone-scrambles: rapid, random sequences of tones drawn from either a major or minor chord. Chapter 1 questions whether low-performing listeners would be able to discriminate between major and minor tone-scrambles if their tones are presented more slowly and with longer duration. Results from this chapter show that the performance of low-performing listeners does not improve when the stimuli are slowed down. In fact, performance was poorest in the slowest experimental condition. This implies that high-performing listeners do not differ from low-performing listeners solely

in being able to extract differences in mode from very rapid stimuli. Chapter 2 investigates whether a 2-hour training regimen in the tone-scramble task is sufficient to improve performance of low-performing listeners. The results show that a small proportion of initially low-performing listeners benefit substantially from training. Most low-performers, however, show no significant improvement. This result shows that the gap that separates low- from high-performing listeners is not easily overcome. Finally, chapter 3 explores whether there are neural differences between the high- and low-performing listeners using fMRI. Results from the chapters of this dissertation suggest that certain listeners possess sensitivity to the difference between major and minor music that others lack and that this difference has little to do with musical training.

## 0.1 Introduction

Musical meaning and emotions have become a facet in music perception that is ubiquitously understood, yet difficult to define. Music engages people of all ages from every cultures in unique patterns of sounds, which is one of the reasons why music makes for such a powerful stimulus. Another reason why music is such a powerful stimulus is due to its multi-dynamic features, such as rhythm, tempo, timbre, or pitch. Yet, these multi-dynamic features also make music perception difficult to study from a foundational level, for it is a combination of these features that ultimately creates the meaning behind what is known as “music” rather than arbitrary patterns of sounds. Numerous studies have looked at the individual components of music to apply to the mechanisms that drive different brain systems underlying actions, emotions, social cognition, and mechanisms of memory and attention (Janata, 2015). However, summarizing the literature constituting all music perception which has crescendoed over the past 20 years would be boundless; therefore, the phenomenon of cognitive and neural differences in tonal perception, specifically to major versus minor mode, is the key focus of this dissertation.

Major and minor musical modes are central to the meaning of music, yet the experience of music is not as simple as the perception of a pure auditory sound. Typically major music sounds “happy” to listeners whereas minor music sounds “sad”. Despite this idea, research has shown that 70% of listeners cannot tell apart the difference when they hear a major versus a minor sequence of notes, but 30% of listeners could, indicating a bimodal distribution in performance across listeners (Chubb et al., 2013; Dean and Chubb, 2017). Many musicians and non-musicians alike scored in the the group that could, but what was surprising about this finding was that many musicians with over 10 years of musical training also fell within the group that could not. Performance in this task correlated with performance in other tasks that required discrimination of other scale-defined qualities besides the major and minor third, suggesting a specialized “scale-sensitivity” that varies dramatically across subjects and that is poorly predicted by the number of years of musical training (Dean and Chubb, 2017). Thus, the experiments of this dissertation build upon this research to investigate whether the bimodal distribution in performance can be eliminated and listeners' performance can improve, and if not, whether this ability to discriminate between the two modes is an inherent trait.

Chapter 1 questioned the speed and complexity of the stimuli used in Chubb et al. (2013), and if the stimuli were to be slowed down and made more musical, would it help the 70% of listeners who performed at chance to hear the difference between the major versus minor stimuli? The short answer is no.

Considering performance could not be improved despite making the stimuli sound closer to music, chapter 2 investigates whether additional training in the task could help performance or enable the low-sensitivity listeners to hear the difference.

Finally, chapter 3 digs deeper into the origin of this bimodal distribution, and whether a difference exists neurally between the high- and low- sensitivity listeners. Specifically, this experiment recruited a subset of high- and low- sensitivity listeners to test with functional magnetic resonance imaging (fMRI) if differences in neural regions or patterns of activity existed between the high- and low- sensitivity listeners.

To conclude, as demonstrated below through the experiments of this dissertation, music is a complex stimulus that modulates behavior in ways that are not generally assumed.

# Chapter 1

# Many listeners cannot discriminate major vs. minor tone-scrambles regardless of presentation rate

## 1.1 Introduction

Music in the major scale sounds "happy" to many listeners whereas music in the minor scale sounds "sad" (Blechner, 1977; Crowder, 1984, 1985a,b; Gagnon and Peretz, 2003; Gerardi and Gerken, 1995; Heinlein, 1928; Hevner, 1935; Kastner and Crowder, 1990; Temperley and Tan, 2013). Because of this striking difference, the major and minor scales have come to play a central role in western music. Surprisingly, however, many listeners have difficulty discriminating major vs. minor melodies (Halpern, 1984; Halpern et al., 1998; Leaver and Halpern, 2004).

Chubb et al. (2013) investigated sensitivity to major vs. minor musical modes using stimuli called tone-scrambles designed to isolate scale-induced perceptual qualities from other aspects of music structure. All tone-scrambles used in their first experiment contained 32, randomly-sequenced, 65ms tones from the Western diatonic scales, including 8 each of the notes $G_5$, $D_6$ and $G_6$ (to establish $G$ as the tonic of each stimulus); in addition, major tone-scrambles contained 8 $B_5$'s (degree 3 of the $G$ major scale) whereas minor tone-scrambles contained 8 $B\flat_5$'s (degree 3 of the $G$ minor scale). On each trial, the listener heard a tone-scramble and strove (with feedback) to classify it as major vs. minor. The distribution of performance was dramatically bimodal: roughly 70% of listeners were near chance; the rest were near perfect.

Evidently, high-performers possess auditory capabilities that low-performers lack. Dean and Chubb (2017) explored the nature of these capabilities. They tested listeners in five tasks. In each, the stimuli were tone-scrambles comprising 32, randomly sequenced, 65ms tones; like the tone-scrambles used by Chubb et al. (2013), all stimuli contained 8 each of the notes $G_5$, $D_6$ and $G_6$ (to establish $G$ as the tonic of every stimulus in all conditions). In addition, each stimulus contained 8 identical target notes. In the "2" task, the target notes were either $A\flat_5$'s or $A_5$'s (lowered second or second scale degree); in the "3" task, the target tones were either $B\flat_5$'s or $B_5$'s (minor or major third scale degree, replicating Chubb et al. (2013)); in the "4" task, the target tones were either $C_6$'s or $D\flat_6$'s (fourth scale degree or tritone); in the "6" task, the target tones were either $E\flat_6$'s or $E_6$'s (minor or major sixth scale degree); and in the "7" task, the target tones were either $F_6$'s or $G\flat_6$'s (minor or major seventh scale degree).

The results were well-described by a "bilinear" model proposing that performance in all five tasks is determined predominantly by a single computational resource $R$. Specifically, for any listener $k$ and any task $t$, let $R_k$ be the amount of $R$ possessed by listener $k$, and let $F_t$ be the strength with which $R$ facilitates task $t$. Then the bilinear model asserts that the sensitivity (as reflected by $d'$) of listener $k$ to the difference between the two types of stimuli in task $t$ is

$$d'_{k,t} = R_k F_t. \tag{1.1}$$

This model accounted for 79% of the variance in $d'_{k,t}$ across 139 listeners and the 5 tasks. Consonant with the results of Chubb et al. (2013), roughly 70% of listeners had $R$-levels near zero, yielding near-chance performance in all 5 tasks; the rest had $R$-levels that enabled much better performance. However, some tasks were facilitated more strongly than others by $R$: the highest levels of performance were achieved in the 2, 3 and 6 tasks; the 4 task was harder, and the 7 task was harder still. Thus $F_2 \approx F_3 \approx F_6 > F_4 > F_7$.

What do these findings suggest about the nature of the resource $R$? The 3, 6 and possibly the 7 tasks require differential sensitivity to the major vs. minor scale. In each of these tasks, one target note belongs to the major but not the minor scale, and the other belongs to the minor but not the major scale. However, in each of the 2 and 4 tasks, one target note belongs to both the major and minor scales, and the other belongs to neither. Dean and Chubb (2017) concluded that $R$ is not specific to the difference between the major vs. minor scales but rather that $R$ confers sensitivity more generally to the spectrum of qualities that music can achieve by creating a scale through establishing a tonic and selecting a subset of notes relative to that tonic. Accordingly, Dean and Chubb (2017) called $R$ "scale-sensitivity."

### 1.1.1 Musical training and scale-sensitivity (SS).

On average, trained musicians have higher SS than non-musicians. In particular, years of musical training and tone-scramble task performance are strongly correlated (Chubb et al., 2013; Dean and Chubb, 2017). Plausibly, however, these correlations are driven by a self-selection bias: listeners high in SS are more likely to seek out musical training than listeners low in SS. This interpretation is supported by the following observations. The positive correlation between years of musical training and SS seen in previous studies (Chubb et al., 2013; Dean and Chubb, 2017) was driven primarily by a large group of listeners with no musical training who had low SS and a smaller group with many years of musical experience who had high SS. Strikingly, however, these studies (Chubb et al., 2013; Dean and Chubb, 2017) have also found a number of listeners with substantial training but low SS as well as other listeners with little training but high SS suggesting that musical training is neither necessary nor sufficient to attain high levels of SS. However, the highest levels of SS seem to be achieved only by listeners with at least 4 years of musical training (Dean and Chubb, 2017). Thus, although musical training is not sufficient for SS, it may be necessary to attain the highest levels of SS.

### 1.1.2 The current study.

Is "scale-sensitivity" really an appropriate name for $R$? The tone-scrambles used in previous studies (Chubb et al., 2013; Dean and Chubb, 2017) all presented tones at the rate of $15.38/sec$ listeners are maximally sensitive to temporal amplitude modulations (of a white noise carrier) from 2 Hz up to 16 Hz with gradually decreasing sensitivity to temporal frequencies above 16 Hz (Viemeister, 1979) suggesting that listeners should be very sensitive to the amplitude modulations use in the tone-scrambles in previous studies (Chubb et al., 2013; Dean and Chubb, 2017).

However, other research has documented strong individual differences in various temporal-sequence-discrimination tasks (Johnson et al., 1987; Kidd et al., 2007) raising the possibility that high-performers in tone-scramble tasks differ from low-performers solely in being able to extract scale-generated qualities from these rapid, musically degenerate stimuli. If so, then if the stimuli are presented more slowly, the gap in sensitivity separating high and low performers should disappear. The main purpose of the current study was to investigate this possibility.

## 1.2 Methods

All methods were approved by the UCI Institutional Review Board.

### 1.2.1 Participants.

Seventy-three UC Irvine undergraduates were compensated with course credit. All had self-reported norrmal hearing. The mean number of years of training was 3.1 (standard deviation: 3.94). Forty-five had at least one year of musical training.

### 1.2.2 Stimuli.

The stimuli tone-scrambles composed of pure tones windowed by a raised cosine function with a 22.5 ms rise time. Five different notes were used, all from the equal-tempered chromatic scale: $G_5$ (783.99 Hz), $B\flat_5$ (932.33 Hz), $B_5$ (987.77 Hz), $D_6$ (1174.66 Hz), and $G_6$ (1567.98 Hz). Table 1 lists the 8 different types of tone-scrambles used. For $n = 1, 2, 4, 8$, the $n$-each stimuli included $n$ each of the notes $G_5$, $D_6$, and $G_6$ as well as $n$ copies of a target note $\mathbf{T}$ which was $B_5$ in $n$-each major stimuli vs. $B\flat_5$ in $n$-each minor stimuli. Each tone in an $n$-each stimulus lasted $\frac{520}{n} ms$. For example, a 4-each-major tone-scramble consisted of a random sequence of 16 tones, each 130ms; 4 of these tones were $G_5$'s, 4 were $B_5$'s, 4 were $D_6$'s and 4 were $G_6$'s. Note that each of the eight different types of tone-scramble had a total duration of 2.08 *sec*.

### 1.2.3 Design.

Listeners were tested in four different, separately blocked tasks: the 1-, 2-, 4- and 8-tasks. On each trial in the $n$-task ($n = 1, 2, 4, 8$), the listener heard an $n$-each stimulus and strove to judge whether it was major or minor. In the response prompt, the two types of stimuli were identified as "HAPPY (major)" and "SAD (minor)"; the listener entered "1" on the keyboard for "major" or "2" for "minor." Correctness feedback was given after each trial, and proportion correct was presented at the end of each block. In each task, the listener performed two blocks of 50 trials with a brief break in between. Each block contained 25 major stimuli and 25 minor stimuli. Task order was counterbalanced across listeners a using Latin square design. For the basic analysis reported in Sec. III, we followed the procedure of the previous tone-scramble studies (Chubb et al., 2013; Dean and Chubb, 2017) and treated the first block of trials as practice and used $d'$ estimated from the second block of 50 trials as our basic dependent measure. (The results would be very similar if $d'$ were computed using all 100 trials.) In the task-specific analyses reported in Sec. V, to increase statistical power, we used all 100 trials from each L.

| task | stimulus type | tone duration | # $G_5$'s | # $B_5\flat$'s | # $B_5$'s | # $D_6$'s | # $G_6$'s |
|---|---|---|---|---|---|---|---|
| 1-task | 1-each-major | 520 ms | 1 | 0 | 1 | 1 | 1 |
| | 1-each-minor | 520 ms | 1 | 1 | 0 | 1 | 1 |
| 2-task | 2-each-major | 260 ms | 2 | 0 | 2 | 2 | 2 |
| | 2-each-minor | 260 ms | 2 | 2 | 0 | 2 | 2 |
| 4-task | 4-each-major | 130 ms | 4 | 0 | 4 | 4 | 4 |
| | 4-each-minor | 130 ms | 4 | 4 | 0 | 4 | 4 |
| 8-task | 8-each-major | 65 ms | 8 | 0 | 8 | 8 | 8 |
| | 8-each-minor | 65 ms | 8 | 8 | 0 | 8 | 8 |

Table 1.1: *The eight different types of tone-scramble used in the current experiment.* For $k = 1, 2, 4, 8$, in the "$k$-task" on each trial the listener heard either a $k$-each-major or a $k$-each-minor tone-scramble and strove to judge which type she had heard. Feedback was given after each trial.

At the start of the experiment, each listener filled out a brief questionnaire. The only information from this questionnaire that is used in the analysis below is the number of years of musical training. Testing was done in a quiet lab on a Windows Dell computer with a standard Realtek audio/sound card using Matlab. Stimuli were presented at the rate of 50000 samples/sec. During testing, the listener wore JBL Elite 300 noise-cancelling headphones with volume adjusted to his or her comfort level. Prior to the first block of trials in the $n$-task (for $n = 1, 2, 4, 8$), the listener was presented with eight, visually identified, example stimuli alternating between $n$-each-major and $n$-each-minor tone-scrambles.

### 1.2.4 Can we account for performance in the 1-, 2-, 4- and 8-tasks in terms of a single cognitive resource?

It is possible that the slower major and minor tone-scrambles used in the 1-, 2- and/or 4-tasks may be discriminable by neural systems other than the system used by high-performing listeners in the 8-task. If so, then multiple cognitive resources may be required to account for variations in performance across all four tasks. In this case, the bilinear model is likely to provide a poor description of the data.

Conversely, if performance is well-described by the bilinear model, then plausibly the same cognitive resource that underlies performance in the 8-task also controls performance in the 1-, 2- and 4-tasks. Note that in this case, if the results for the 8-task replicate previous findings (Chubb et al., 2013; Dean and Chubb, 2017), then the majority of listeners $k$ will perform poorly in the 8-task, implying that they have levels of $R_k$ near 0, implying in turn that they will also perform poorly in the 1-, 2- and 4-tasks.

In the current context, the bilinear model proposes that the values $d'_{k,t}$ achieved by all 73 listeners $k$ in all four tasks $t$ are captured by Eq. 1.1, where $R_k$ is the amount of $R$ possessed by listener $k$ and $F_t$ is the strength with which task $t$ is facilitated by $R$.

Note that the model of Eq. 1.1 is underconstrained. We get exactly the same predictions from a model in which (1) $R_k$ is rescaled by an arbitrary factor $A$ and (2) $F_t$ is rescaled by $A^{-1}$. To avoid this problem, we impose the constraint that

$$\sum_{\text{tasks } t} F_t = 4 \qquad \text{(where 4 is the number of tasks).} \tag{1.2}$$

Eq. 1.2 makes it easy to interpret the results. If all four tasks are equally facilitated by $R$, then $F_t$ will be 1 for all tasks. Deviations from 1 directly indicate relative facilitation strength. In addition, imposing this constraint also makes $R_k$ a prediction of the average value of $d'_{k,t}$ achieved by listener $k$ across the four tasks. (Note, however, that Dean and Chubb (2017) imposed a different constraint: they forced the sum of squared $F_t$ values to be 1.)

## 1.3 Results

In computing $d'$ values for the analyses reported in this section, we used only the last 50 trials in each task (treating the first 50 as practice). If a listener responded correctly on all 25 "major" (or "minor") stimuli, the probability of a correct "major" ("minor") response was adjusted to $\frac{24.5}{25} = 0.98$ (Macmillan and Kaplan, 1985). This yields $d' = 4.1075$ for listeners who perform perfectly across all 50 trials.

For each pair of tasks, Fig. 1.1 plots the $d'$s achieved (by all listeners) in one task against those achieved in the other. Note first that in each scatterplot, there is (i) a large group of listeners for whom $d'$ is near 0 in both tasks and (ii) a smaller group of listeners who achieve levels of $d'$ near 4.1075 in both tasks. In addition, there are other listeners intermediate between these two extreme groups. The relative difficulty of the two tasks being compared in a given scatterplot is evident in the distribution of $d'$s for this intermediate group.

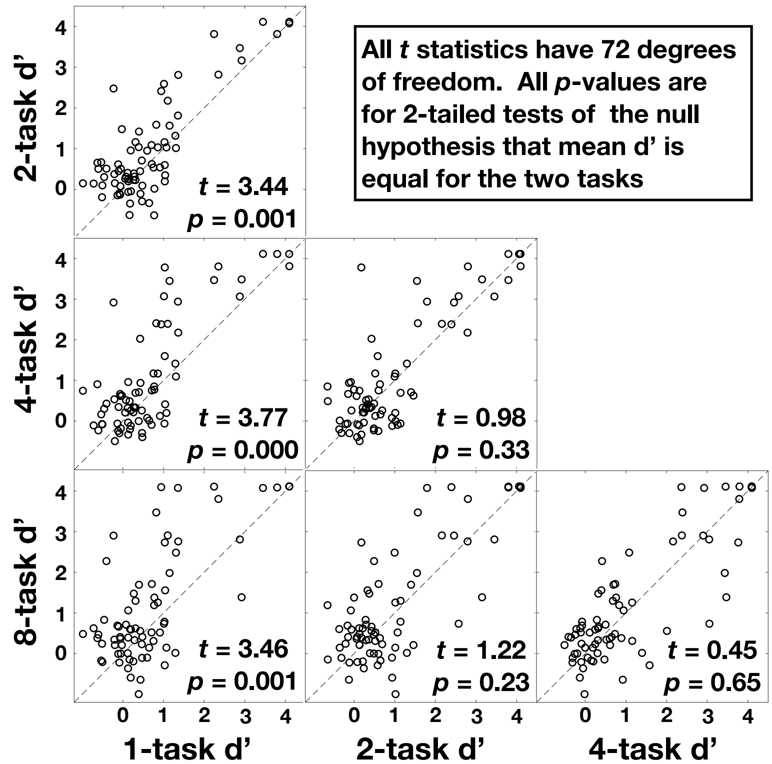


Figure 1.1: *Scatterplots of $d'$ achieved by all Ls in each pair of tasks.* $d'$ estimates are based on the last 50 trials in each task. Associated with each scatterplot are a $t$-statistic and the corresponding $p$-value derived from a paired samples $t$-test of the null hypothesis that the mean value of $d'$ is equal for the two tasks. Note that the 1-task yields lower values of $d'$ than the all of the other tasks.

In each scatterplot in the left column of Fig. 1.1, the points corresponding to these intermediate listeners fall above the main diagonal; this reveals that the intermediate listeners tend to achieve higher levels of $d'$ in each of the 2-, 4- and 8-tasks than they do in the 1-task.

That these differences in task difficulty are significant is confirmed by paired-samples $t$-tests (null hypothesis: the mean values of $d'$ are equal for the tasks being compared); the $p$-values comparing the 1-task to the 2-, 4- and 8-tasks are all highly significant. By contrast, none of the tests comparing the mean $d'$s between the 2-, 4- and 8-tasks reach significance. We conclude that the 1-task is more difficult than the other three tasks, which are roughly equal in difficulty.

### 1.3.1 The bilinear model results.

The estimated values of $F_t$ are shown in Fig. 1.2. Consonant with the $t$-test results described above, $F_{\text{1-task}}$ is lower than all of $F_{\text{2-task}}$, $F_{\text{4-task}}$ and $F_{\text{8-task}}$. Below we address the question of why this is.

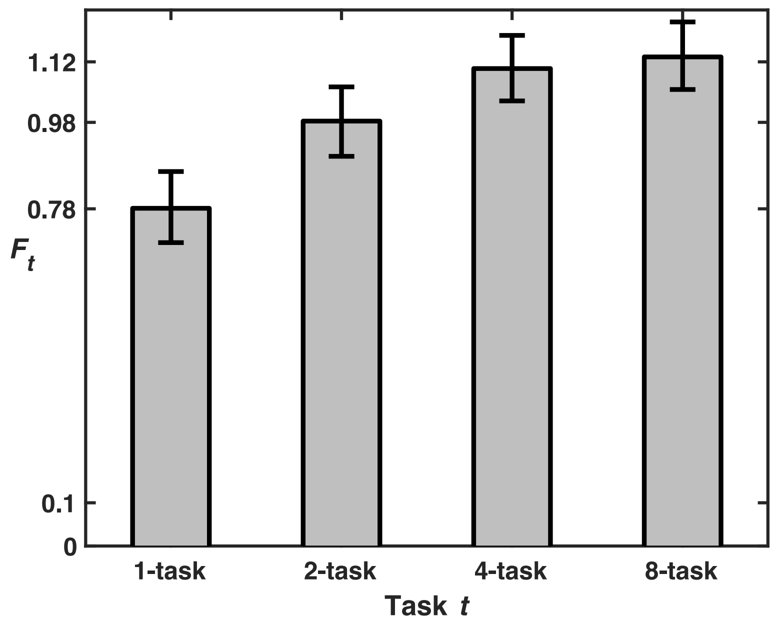


Figure 1.2: *Estimated values of $F_t$ for the four tasks.* As suggested by Fig. 1.1, $F_{\text{1-task}}$ is lower than all of $F_{\text{2-task}}$, $F_{\text{4-task}}$ and $F_{\text{8-task}}$, reflecting the fact that performance in the 1-task is facilitated less strongly by the cognitive resource $R$ than are the other three tasks. Error bars are 95% Bayesian credible intervals.

The left panel of Figure 1.3 shows the histogram of $R$ levels estimated for our 73 listeners. As found by Dean and Chubb (2017), the histogram is positively skewed with many listeners with $R$ values near 0 but does not appear bimodal. However, the histogram of predicted proportions correct in the 8-task (assuming listeners used optimal criteria) is bimodal and similar to that observed in Experiment 1 of Chubb et al. (2013).

The bilinear model provides an excellent description of the results. This is shown by Fig. 1.4 which plots the estimates of $d'_{k,t}$ derived individually from the the data for each listener $k$ in each task $t$ against the values predicted by the bilinear model. The bilinear model accounts

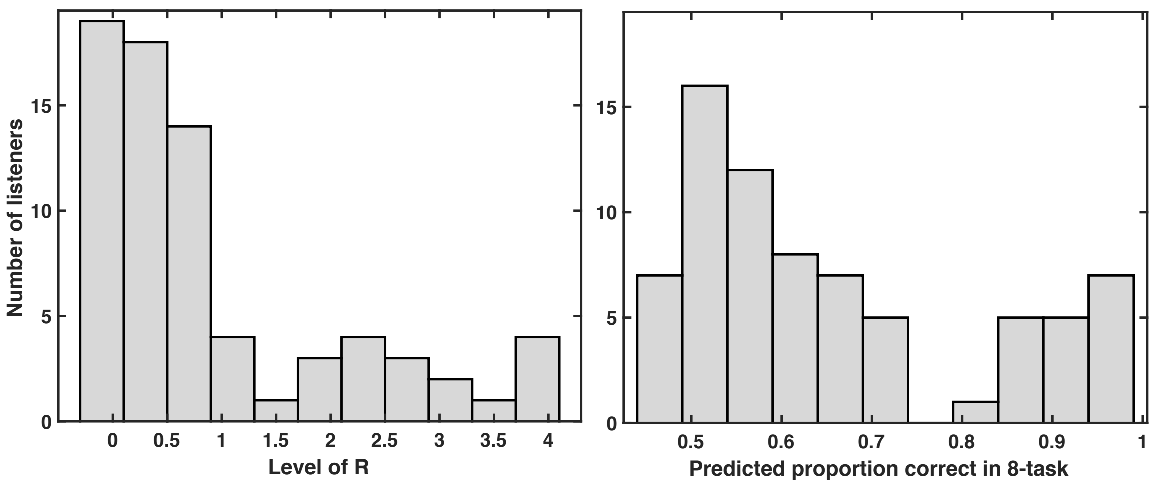


Figure 1.3: Left panel: Histogram of estimated $R$ levels for all Ls. This histogram shows high skew, with a large number of Ls with $R$ values near 0 and a long positive tail. Right panel: the histogram of predicted proportion correct in the 8-task corresponding to the $R$ levels in the left panel. This histogram appears bimodal even though the histogram of $R$ levels does not.

for 84% of the variance in the values of $d'_{k,t}$ for our 73 listeners across the 1-, 2-, 4- and 8-tasks, reflecting the strong relationship visible in Fig. 1.4.

However, the bilinear model does not account for all of the structure in the data. An $F$-test comparing the bilinear model with a model allowing $d'_{k,t}$ to depend two cognitive resources yields $F(74, 142) = 1.93$, $p = 0.0004$ implying that the 2-resource model fits significantly better. Adding a third cognitive resource does not significantly improve the fit with two resources ($F(72, 70) = 1.17$, $p = 0.25$). We conclude that performance is influenced slightly but significantly by variations in a single cognitive resource in addition to $R$.

### 1.3.2 The effect of music training.

Fig. 1.5 plots listeners' $R$-levels as a function of their years of musical training. The correlation of 0.50 is due mainly to many listeners with no musical training who have low levels of $R$. Of the 25 listeners $k$ in our sample who had 5 or more years of musical training, 13 had $R_k < 1$ (corresponding to $< 70\%$ correct across all four tasks). Note, in particular, that

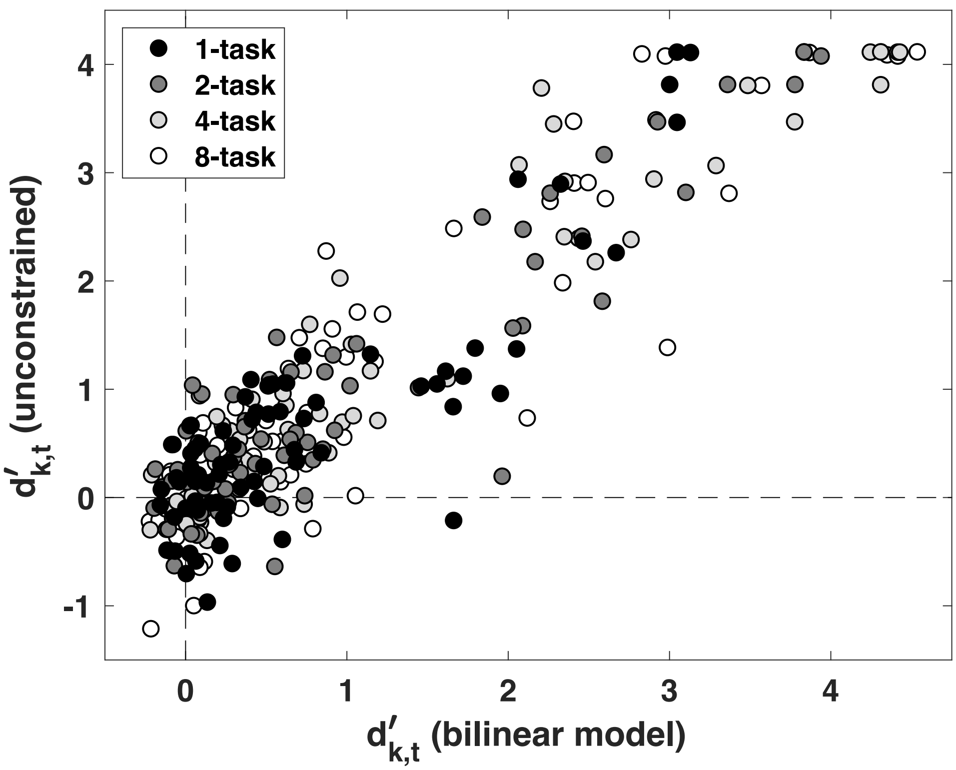


Figure 1.4: *The description provided by the bilinear model.* This scatterplot contains a point for each listener $k$ in each task $t$. The abscissa is the value of $d'_{k,t}$ predicted by the bilinear model, and the ordinate is the corresponding maximum likelihood estimate of $d'_{k,t}$ derived independently from the the data for listener $k$ in task $t$. As indicated by the legend, points derived from different tasks $t$ are plotted in different gray-scales.

the lone listener with 18 years of musical experience had SS 0.32. Many such listeners with substantial musical training but low levels of $R$ have been previously reported (Chubb et al., 2013; Dean and Chubb, 2017).

Strikingly, however, in both the study of Dean and Chubb (2017) and the current study, listeners with few years of musical training invariably possess lower levels of $R$ than do the most sensitive listeners with five or more years of musical training. Out of all of the 43 listeners, $k$, in the current sample with two or fewer years of musical training, none achieved levels of $R_k$ higher than around 2 (corresponding to an average proportion correct around 86% across all four tasks). By contrast, four of the 25 listeners $k$ with 5 or more years of musical training had $R_k$ near 4 (corresponding to near perfect performance across all four tasks).

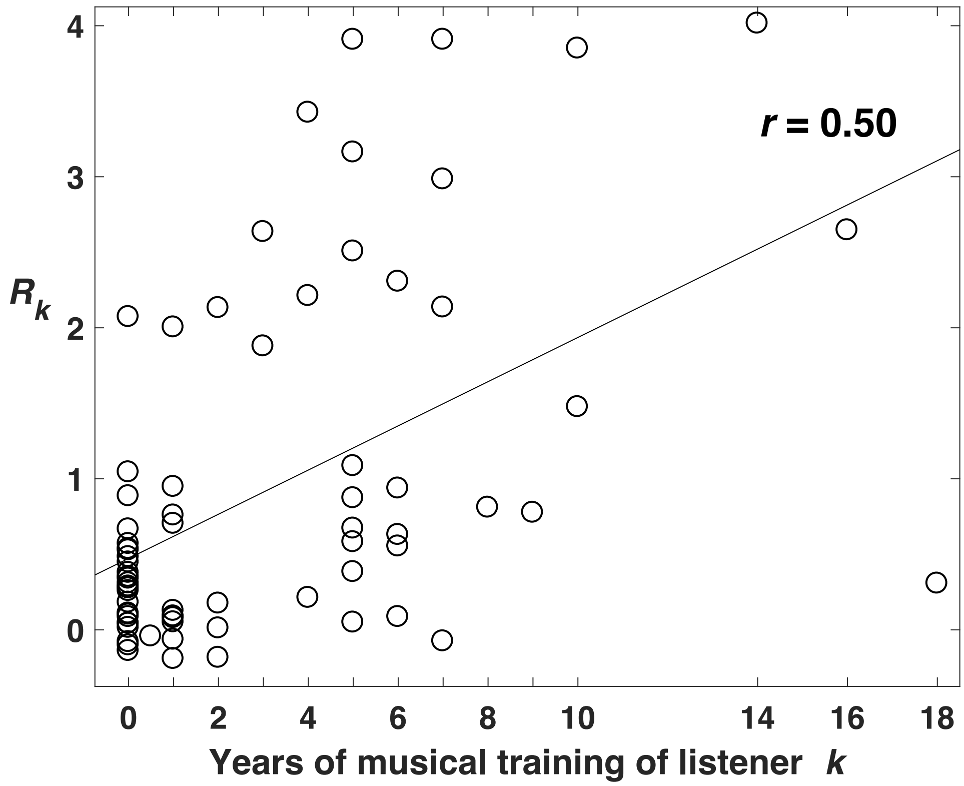


Figure 1.5: *The relation between years of musical training and R.*

## 1.4 Discussion of the present study

The current study sought to determine whether low-performers in tone-scramble tasks (Chubb et al., 2013; Dean and Chubb, 2017) would do better if the stimuli were presented more slowly. The answer is no. In particular, the 1-task – in which each note was presented for more than half a second – proved to be the most difficult.

Moreover, performance across all of the 1-, 2-, 4- and 8-tasks is well-described by the bilinear model of Eq. 1.1 suggesting that performance in these four tasks depends predominantly on a single cognitive resource. Given that the 8-task was included both in the current study and in the study of Dean and Chubb (2017), it is likely that a single cognitive resource $R$

controls performance in all the tasks used in the current study and the study of Dean and Chubb (2017).

In ruling out the possibility that high-performers in the 8-task differ from low-performers merely in their ability to extract scale-defined qualities from very rapid note sequences, the current findings bolster the interpretation of $R$ suggested by the name "SS," i.e., that $R$ confers general sensitivity to the range of qualities that music can achieve by being in a scale (Dean and Chubb, 2017).

A final question remains from the current study. This is the focus of the next section.

## 1.5 What makes the 1-task harder than the 2-, 4- and 8-tasks?

### 1.5.1 Some notation.

It will be convenient to indicate the notes $G_5$, $B\flat_5$, $B_5$, $D_6$ and $G_6$ by the integers $1, 4, 5, 8$, and 13, reflecting their locations in the chromatic scale starting at $G_5$. We will refer to these numbers as the "pitch-heights" of notes. We shall also use the symbol "**T**" to refer to the "target" note (4 or 5) in a given stimulus. Thus, if a stimulus $S$ has $\mathbf{T} = 4$, then $S$ is minor; if $S$ has $\mathbf{T} = 5$, then $S$ is major. A "note-order" is a permutation of the four symbols, "1", "8", "13" and "**T**"; by substituting "4" ("5") for **T**, we obtain a symbol string corresponding to a minor (major) stimulus. Finally, for a given note-order $Q$, we will write $S_Q^+$ for the stimulus with note-order $Q$ that has $\mathbf{T} = 5$ and $S_Q^-$ for the stimulus with note-order $Q$ that has $\mathbf{T} = 4$.

### 1.5.2 The order of the tones in a stimulus affects responding.

As revealed by Fig. 1.6, the results from the 1-task show unanticipated structure. This figure presents results averaged across all 100 trials performed by each of our 73 listeners. For each stimulus $S$, Fig. 1.6 plots the proportion of all of the trials on which $S$ was presented that the response was "major." If the judgments of our listeners depended only on whether $S$ has $\mathbf{T} = 4$ or $\mathbf{T} = 5$, then across the 24 different note-orders on the horizontal axis, each of the plots for $S$ with $\mathbf{T} = 5$ and $\mathbf{T} = 4$ would be flat. On the contrary, we see dramatic and roughly parallel variations in the two curves suggesting that, regardless of whether the stimulus is major or minor, the order of the notes strongly influences our listeners' responses.

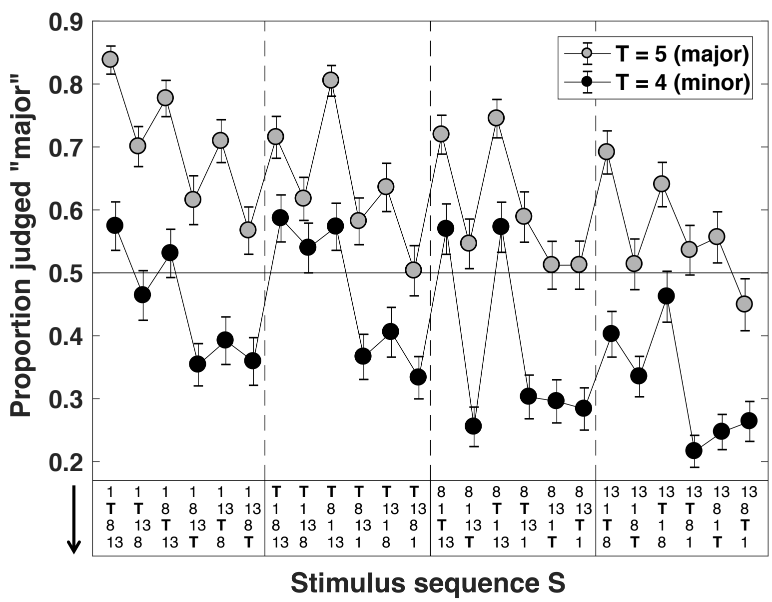


Figure 1.6: *The proportion of "major" responses to all 48 task-1 stimuli $S$.* The plot pools the data across all 100 trials performed by all 73 Ls. The gray markers show the proportion of "major" responses for major stimuli (target note $\mathbf{T} = 5$); the black markers show the proportion of "major" responses for the corresponding minor stimuli (target note $\mathbf{T} = 4$). The note-order of a given stimulus $S$ is shown (running downward) at the bottom of the figure; $1 = G_5$, $4 = B\flat_5$, $5 = B_5$, $8 = D_6$, $13 = G_6$. Error bars are 95% confidence intervals for the mean proportion.

### 1.5.3 Modeling note-order effects.

Note that listeners $k$ whose levels of $R_k$ are very high are likely to perform perfectly in the 1-task; because the responses of such listeners depend only on whether $\mathbf{T} = 4$ or $\mathbf{T} = 5$, they will necessarily be invariant with respect to the order of the notes in the stimulus. Conversely, the responses of any listeners whose level of $R_k$ is insufficient to produce perfect performance may show dependencies on the order of notes in the stimulus. It is difficult, however, to anticipate the pattern of such dependencies. For this reason, both of the models described below allow the strength of note-order effects to vary freely as a function of $R_k$.

#### 1.5.3.1 The general modeling framework.

Both of the models we consider assume that when listener $k$ is presented with stimulus $S$, the listener computes a noisy internal statistic and compares it to a criterion $\eta_k$ that is fixed across all trials performed in the 1-task by listener $k$. If this internal statistic is larger than $\eta_k$, the listener responds "major"; otherwise, the listener responds "minor." We use the symbol "$M_{k,S}$" to denote the expectation of this internal statistic, and we assume that the noise is Gaussian with standard deviation 1. Formally,

$$\text{Response of L } k \text{ to stimulus } S = \begin{cases} \text{“Major”} & \text{if } M_{k,S} + X > \eta_k \\ \text{“Minor”} & \text{if } M_{k,S} + X < \eta_k \end{cases}, \qquad (1.3)$$

where $X$ is a standard normal random variable.

The fullest model of the form expressed by Eq. 1.3 allows $M_{k,S}$ to be a free parameter for each listener $k$ and each task $S$. However, this model provides no traction in understanding the effects revealed by Fig. 1.6. Instead, we consider two nested models below, each of which is nested within the fullest model.

The first model we consider is called the "descriptive" model because its purpose is to describe with full freedom the variations in the stimulus-specific biases present in the data. Accordingly, the descriptive model includes free parameters for all of these biases. The "pitch-height-biased" model described below will attempt to capture with fewer parameters the pattern of the biases revealed by the descriptive model.

### 1.5.4 Fitting procedures.

To estimate parameter values for each of the descriptive and pitch-height-biased models, we use a Bayesian fitting procedure. Specifically, we assume a jointly uniform prior distribution with wide bounds on all model parameters. We then use Markov chain Monte Carlo simulation to extract a sample of vectors from the posterior joint density characterizing the parameters. All of the error bars in figures plotting parameter values give the 0.025 and 0.975 quantiles of the marginal density for the given parameter.

### 1.5.5 The descriptive model.

For any stimulus $S$, let

$$\tau_S = \begin{cases} 1 & \text{if stimulus } S \text{ has } \mathbf{T} = 5, \\ -1 & \text{if stimulus } S \text{ has } \mathbf{T} = 4. \end{cases} \tag{1.4}$$

Thus, $\tau_S$ is the variable that the listener *should* be using to make her judgment on each trial. However, Fig. 1.6 suggests that different stimuli $S$ inject stimulus-specific biases into the judgments of many listeners $k$. For any stimulus $S$, we write $\beta_S$ for the bias associated with $S$. We assume that the influence on $M_{k,S}$ both of $\tau_S$ and also of $\beta_S$ depends on $R_k$.

Specifically, the descriptive model assumes that

$$M_{k,S} = f_\tau(R_k)\tau_S + f_\beta(R_k)\beta_S, \tag{1.5}$$

where the function $f_\tau(R)$ reflects the strength with which the response of a listener with SS $R$ is influenced by $\tau_S$ (i.e., whether $S$ has $\mathbf{T} = 4$ or $\mathbf{T} = 5$), and the function $f_\beta(R)$ reflects the strength with which the response of a listener with SS $R$ is influenced by the stimulus-specific bias $\beta_S$.

To limit the number of parameters contributed to the model by $f_\tau(R)$ and $f_\beta(R)$, we force each of these functions to assign a fixed value to all $R_k$ in a given sextile of the distribution of scale-sensitivities observed across all listeners $k$ in the current study.

In order for the descriptive model (Eq. 1.5) to be uniquely specified, additional constraints need to be imposed on the stimulus-specific biases $\beta_S$. Specifically, we require that

$$\sum_S \beta_S = 0 \qquad \text{and} \qquad \frac{1}{48}\sum_S \beta_S^2 = 1, \tag{1.6}$$

where each sum is over all 48 stimuli $S$. The first constraint insures that the $\beta_S$ values can't trade off with the threshold values $\eta_k$. The second constraint insures that the $\beta_S$ values can't trade off with $f_\beta$. The specific form of this second constraint on the $\beta_S$ values is chosen to make their magnitudes comparable to those of the $\tau_S$ values (which also satisfy $\frac{1}{48}\sum_S \tau_S^2 = 1$).

The parameters of the descriptive model are the 48 $\beta_S$ values, the six values of each of $f_\tau$ and $f_\beta$ and the 73 values of $\eta_k$. Therefore, taking into account the two degrees of freedom sacrificed by imposing the constraints of Eq. 1.6 on the $\beta_S$ values, the model has 73+48+12-2 = 131 degrees of freedom.

The main aim of this model is to derive estimates of the stimulus-specific biases, $\beta_S$, that are cleanly dissociated from the powerful influences that we know are exerted by $\tau_S$ on the responses of listeners high in SS. We naturally expect $f_\tau(R)$ to increase strongly with $R$. We also know that for very high values of $R$, $f_\beta(R)$ must tend to 0; however, we have no strong *a priori* expectations concerning the form of $f_\tau(R)$ for lower values of $R$.

### 1.5.6 Results from the descriptive model

The stimulus-specific biases $\beta_S$ for all 48 stimuli $S$ are shown in Fig. 1.7. Across the 24 note-orders on the horizontal axis, the gray markers show the results for stimuli $S$ with $\mathbf{T}$ = 5, and the black markers show the results for $S$ with $\mathbf{T}$ = 4. The main thing to note is that there are dramatic differences between the stimulus-specific biases for different stimuli, and these biases appear similar for major and minor stimuli with the same note-order.

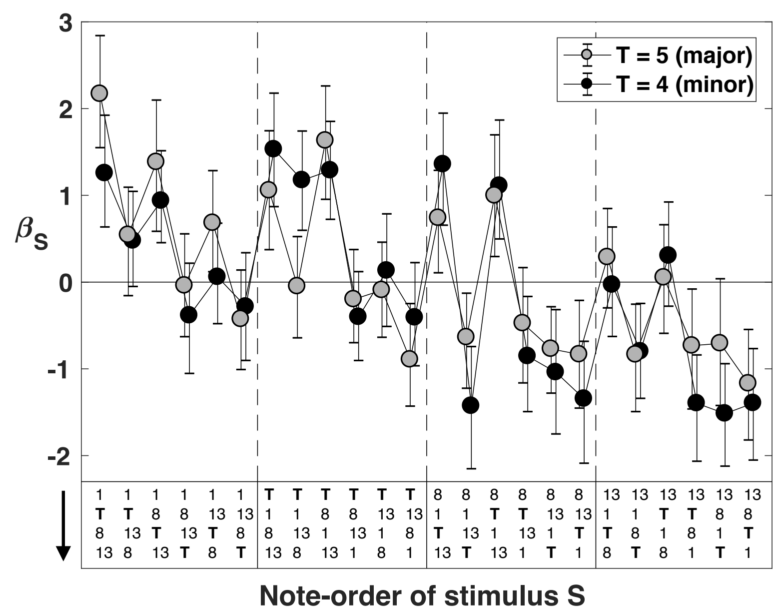


Figure 1.7: *The stimulus-specific biases $\beta_S$ for the 48 1-task stimuli $S$.* The gray markers show the stimulus-specific biases for stimuli $S$ with $\mathbf{T}$ = 5; the black markers show the stimulus-specific biases for stimuli $S$ with $\mathbf{T}$ = 4. The note-order of a given stimulus $S$ is shown (running downward) at the bottom of the figure. Error bars are 95% Bayesian credible intervals.

It is useful to replot the data in terms of the following two statistics:

$$\delta_Q = \frac{\beta_{S_Q^+} - \beta_{S_Q^-}}{2} \tag{1.7}$$

and

$$\mu_Q = \frac{\beta_{S_Q^+} + \beta_{S_Q^-}}{2}. \tag{1.8}$$

Thus $\delta_Q$ reflects sensitivity to the difference between target notes 5 vs. 4 in the context of note-order $Q$, and $\mu_Q$ reflects the bias injected by note-order $Q$ regardless of the target note.

The black disks in the left (right) panel of Fig. 1.8 plot $\mu_Q$ ($\delta_Q$) for all 24 note-orders. The gray triangles in the left panel plot the fit provided by the nested "pitch-height-biased" model described below. The fact that nearly all of the credible intervals in the right panel contain 0 shows that note-order exerts little or no influence on sensitivity to the difference between target notes 5 vs. 4. By contrast, as shown by the black disks in the left panel, the note-order-specific bias $\mu_Q$ depends strongly on $Q$.

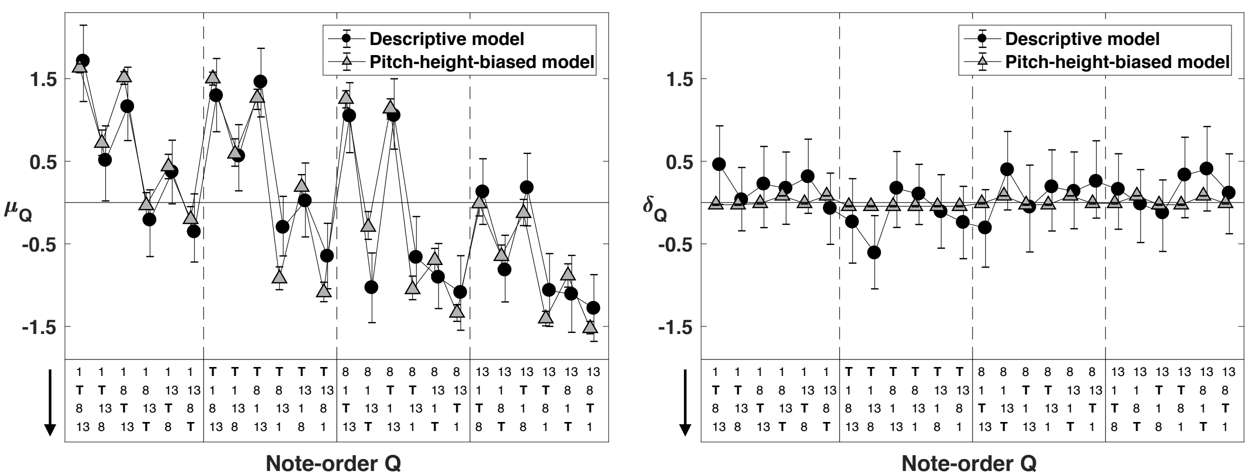


Figure 1.8: *The note-order-specific biases $\mu_Q$ (left panel) and note-order-specific sensitivities $\delta_Q$ (right panel) for the 24 note-orders $Q$.* The note-orders on the horizontal axis run downward. The black circles show the estimates derived from the descriptive model in which all stimulus-specific biases are free parameters. The gray triangles show the predictions of the pitch-height-biased model. Error bars are 95% Bayesian credible intervals.

The upper left panel of Fig. 1.10 plots the functions $f_\tau$ (black) and $f_\beta$ (gray) in the 1-task. As expected, $f_\tau$ increases with with $R_k$. By contrast, $f_\beta$ remains flat (and significantly greater than 0) across all 6 sextiles of the the distribution of $R_k$ (with a 50% drop-off for the last sextile). This implies that whatever factors are operating to produce the note-order-specific biases $\mu_Q$ are largely invariant with respect to $R_k$.

### 1.5.7 The pitch-height-biased model.

Mean pitch-height has been shown to exert strong influence on performance in classifying major vs. minor tone-scrambles; in particular, tone-scrambles that include more high tonics ($G_6$'s) than low tonics ($G_5$'s) are more likely to be judged major (Chubb et al., 2013). This finding suggests that the variations in $\mu_Q$ (the black disks in the left panel of Fig. 1.8) might depend on the variations in pitch-height across different notes in the stimulus.

The pitch-height-biased model expresses this intuition. Writing $S(t)$ for the pitch-height of the note in temporal location $t = 1, 2, 3, 4$, we estimate the stimulus-specific bias $\beta_S$ of a given stimulus $S$ as follows:

$$\beta_S = \sum_{t=1}^{4} w_t \left( S(t) - 6.625 \right) \tag{1.9}$$

where the $w_t$'s are model parameters called "sequential pitch-height weights," and the constant 6.625 is the expected pitch-height of the note occurring on any given trial at any one of the four sequence locations in the 1-task. The four sequential pitch-height weights actually use only two degrees of freedom because, in order for the model to be uniquely determined, the $w_t$'s must (1) sum to 0, and (2) be chosen so that the resulting $\beta_S$'s satisfy the right side of Eq. 1.6. The total number of degrees of freedom in the pitch-height-biased model is 87: 2 degrees of freedom for the 48 $\beta_S$ parameters, 73 for the $\eta_k$'s, and 6 for each of the functions $f_J$ and $f_\beta$.

### 1.5.8 Results from the pitch-height-biased model: the importance of ending on a high note.

The data are captured very cleanly by the pitch-height-biased model. This is shown by Fig. 1.8. The left (right) panel of Fig. 1.8 plots the $\mu_Q$ ($\delta_Q$) values given by the descriptive model (black disks) along with the estimates given by the pitch-height-biased model (gray triangles). The patterns are strikingly similar even though the pitch-height-biased model uses only two degrees of freedom to estimate all of the $\mu_Q$ and $\delta_Q$ values of the descriptive model (48 parameters that take 46 degrees of freedom).

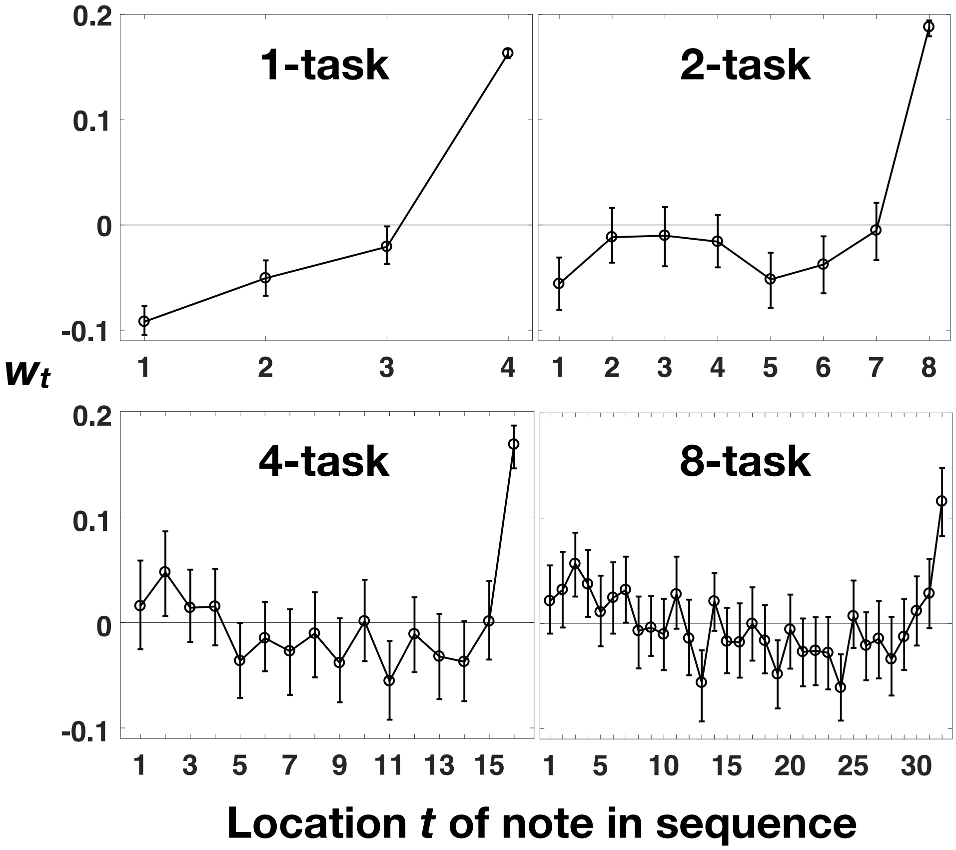


Figure 1.9: *The pitch-height weights for all four tasks.* The sequential pitch-height weights $w_t$, for all locations $t$ in the stimulus in each of the four tasks. These weights are used in the pitch-heigh-biased model (Eq. 1.9) to estimate the values $\beta_S$ for all 48 stimuli $S$, and consequently also the values $\mu_Q$ and $\delta_Q$ for all note-orders $Q$. Error bars are 95% Bayesian credible intervals.

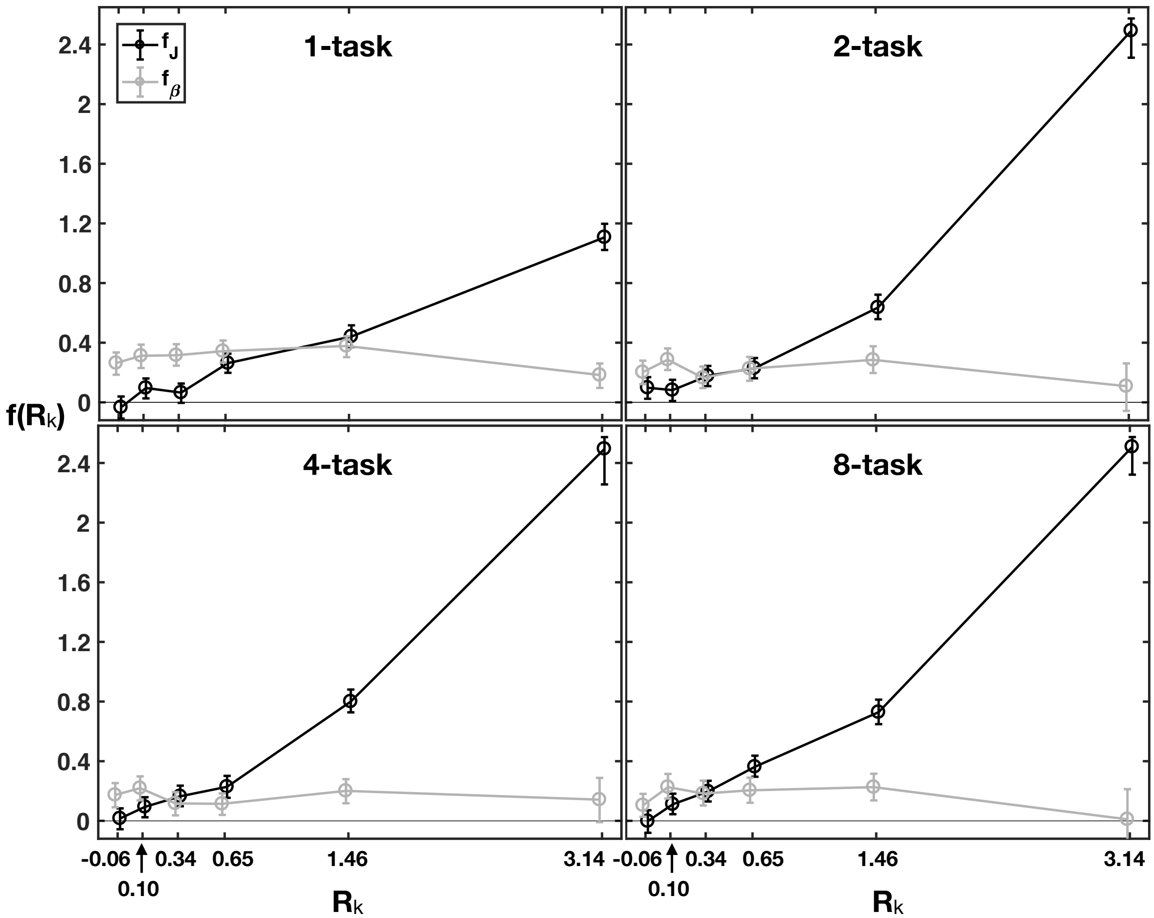


Figure 1.10: *The functions $f_\tau$ and $f_\beta$ in all four tasks.* The values of $R_k$ plotted on the horizontal axis are the mean values of the six sextiles of $R_k$ observed across the 73 Ls $k$ tested. The black (gray) plot shows $f_\tau(R_k)$ ($f_\beta(R_k)$). Error bars are 95% Bayesian credible intervals.

The four pitch-height weights used in Eq. 1.9 to estimate the $\beta_S$ values predicted by the pitch-height-biased model (and consequently also the $\mu_Q$ and $\delta_Q$ values plotted by the gray triangles in Fig. 1.8) are plotted in the upper left panel of Fig. 1.9. This panel shows that the bias injected by a particular note-order $Q$ in the 1-task depends primarily on the pitch-height of the last note (although the weights of the first three notes also significantly influence mean bias). If the note-order ends on a high note, then the bias will be toward a "major" response.

### 1.5.9 Do the 2-, 4- and 8-tasks also show the ending-on-a-high-note Effect?

Although the 2-, 4-, and 8-tasks include too many possible stimuli to allow us to fit the descriptive model in which the influence of each note-order is reflected by a free parameter, we can easily fit the pitch-height-biased model to the data from each of these conditions. The resulting sequential pitch-height weights $w_t$ for the 2-, 4- and 8-tasks are shown in the upper right, lower left and lower right panels of Fig. 1.9. Each of these functions shows that ending on a high note introduces a bias to respond "major." The upper right, lower left and lower right panels of Fig. 1.10 show the $f_\tau$ and $f_\beta$ functions for these three tasks. The relative influence of note-order is lower for the 2-, 4- and 8-tasks than it is for the 1-task. However, note-order exerts a significant influence on the responses of all listeners $k$ other than those with levels of $R_k$ in highest sextile (who perform nearly perfectly in each of the 2-, 4- and 8-tasks).

### 1.5.10 Discussion of the note-order effects.

The note-order effects revealed by the analysis are unanticipated. Note-order is entirely irrelevant to the task; thus, the trial-by-trial feedback that listeners received throughout this experiment should suppress effects of this sort. Nonetheless, in all four tasks, our listeners show a powerful bias to respond "major" to stimuli that end on a high note.

The ending-on-a-high-note bias seems to be unrelated to scale sensitivity. The flatness of the gray curve in the upper left panel of Fig. 1.10 shows that this bias operates in the 1-task with nearly equal strength across all listeners irrespective of their scale-sensitivities. Similar effects are seen in the other panels of Fig. 1.10, except that in the 2-, 4- and 8-tasks listeners $k$ with $R_k$ values in the highest sextile are largely immune to the ending-on-a-high-note bias.

We speculate that our instructions to classify stimuli as "HAPPY (major)" or "SAD (minor)" may have produced the ending-on-a-high-note effects. Although the current study sheds no light on exactly why ending on high (low) note has the effect of making a tone-scramble sound "HAPPY" ("SAD"), theories about the relationship between music and speech may shed some light into this effect (Patel, 2005; Patel et al., 2006). In particular, when a speaker asks a question, or is excited, elated, or happy, the intonation of the voice is steered toward ending on a high pitch (Juslin and Laukka, 2003; Swaminathan and Schellenberg, 2015; Curtis and Bharucha, 2010). Thus, if music inherits some of its emotional expressiveness from our sensitivity to speech variations, then we might expect the ending-on-a-high-note effect (Romano, 2002).

## 1.6 General discussion

The current results support the idea that performance of a listener in all of the tasks tested in the current study as well as those tested by Dean and Chubb (2017) is determined by the listener's level of a single cognitive resource $R$. The current results bolster the proposal of Dean and Chubb (2017) that $R$ confers general sensitivity to the qualities that music can achieve by establishing a tonic and selecting a set of notes relative to the tonic to serve as a scale. This proposal raises the following basic question.

### 1.6.1 How should we think about scale-defined qualities? The analogy to color.

The main purpose of this section is to clarify what we take to be the central open questions concerning SS. We will first describe a working hypothesis concerning the nature of SS. We will then state the main questions raised in light of this hypothesis. Finally, we will discuss possible methods for addressing these questions.

The qualities of many auditory textures seem to be represented by additive summary statistics (McDermott and Simoncelli, 2011; McDermott et al., 2013), i.e., by neural processes that compute the average of some temporally local statistic over an extended time window. Plausibly, this is the case for the qualities that tone-scrambles can evoke. Moreover, the fact that the performance of listeners with high levels of SS is unperturbed by the random sequencing of tone-scrambles suggests that the summary statistics corresponding to scale-defined qualities do not depend on note-order but only on the proportions of different notes in the stimulus–i.e., on the scale of the stimulus, where the term "scale" refers to the histogram of notes in the stimulus.

These observations suggest that the auditory system may encode scale-defined qualities analogously to the way the visual system encodes colors. Note in particular that just as the scale-defined quality evoked by a tone-scramble is determined by the proportions of different notes it contains (i.e., by its scale), the color of a light of some fixed quantal flux is similarly determined by the proportions of different-wavelength quanta it contains (i.e., by its spectrum).

The color of a light for photopic vision is determined by three summary statistics: the activations produced by the light in the $L$-, $M$- and $S$-cone classes (i.e., the long-, medium- and short-wavelength sensitive cone classes). For example,

$$L\text{-cone activation produced by light } H = K_H \sum f_L(\lambda) P_H(\lambda) \tag{1.10}$$

where the sum is over all wavelengths $\lambda$ that occur in the light, $K_H$ is a constant that depends on the quantal flux of $H$, $P_H(\lambda)$ gives the proportion of quanta in $H$ that have wavelength $\lambda$, and the function $f_L(\lambda)$ gives the sensitivity of $L$-cones to quanta of wavelength $\lambda$.

If scale-defined qualities are analogous to colors, then for some set of scale-sensitive neuron classes $C_1, C_2, \cdots, C_N$, the scale-defined quality evoked by a tone-scramble is determined by the activations that the tone-scramble produces in these $N$ neuron classes. For a given neuron class, $C_i$,

$$\text{Time-averaged activation produced in } C_i \text{ by a tone-scramble } S = K_S \sum f_i(d) P_S(d) \tag{1.11}$$

where the sum is over all scale-degrees $d$ (relative to whichever note has been established as the tonic) that occur in the tone-scramble $S$, $K_S$ is a constant that depends on the rate of presentation of the tones in $S$, $P_S(d)$ gives the proportion of tones in $S$ that have scale-degree $d$, and $f_i(d)$ gives the sensitivity of neurons in class $C_i$ to tones of scale-degree $d$.

Of course actual melodies differ from tone-scrambles in that the notes they contain may vary in such ancillary qualities as attack, timbre, duration, loudness and rhythmic context. Plausibly, however, these complexities can be handled with a simple modification of Eq. 1.11. Let $S(t)$ be the scale degree of the $t^{th}$ tone in tone-scramble $S$, and let $N_S$ be the total number of tones in $S$. Then we can rewrite Eq. 1.11 as follows:

$$\text{Time-averaged activation produced in } C_i \text{ by a tone-scramble } S = \frac{1}{N_S} \sum_{t=1}^{N_S} f_i(S(t)). \tag{1.12}$$

If we suppose that the ancillary features of a given note in a melody $M$ collectively operate only to modify the weight exerted by that note in activating a given neuron class $C_i$, then we can generalize Eq. 1.12 as follows:

$$\text{Time-averaged activation produced in } C_i \text{ by melody } M = \frac{1}{\text{Total } M\text{-weight}} \sum_{t=1}^{N_M} f_i(M(t)) W_M(t) \tag{1.13}$$

where $M(t)$ is the scale-degree of the $t^{th}$ note of $M$, $N_M$ is the number of notes in $M$, and $W_M(t)$ is the weight (determined by the ancillary features of the $t^{th}$ note of $M$) exerted by the $t^{th}$ note in $M$, and

$$\text{Total } M\text{-weight} = \sum_{t=1}^{N_M} W_M(t). \tag{1.14}$$

In summary, if the auditory system represents scale-defined qualities analogously to the way in which the visual system represents colors, then (for listeners with non-zero SS)

1. The auditory system contains neuron classes $C_i$, $i = 1, 2, \cdots, N$, each of which confers sensitivity to a different dimension of scale-defined quality.

2. The scale-defined quality sensed by neurons in class $C_i$ is characterized by a sensitivity function $f_i(d)$ that reflects the influence exerted on the activation of neurons in this class by tones of scale degree $d$ (relative to whichever note has been established as the tonic).

3. The time-averaged activation produced in a neuron in class $C_i$ by a tone-scramble $S$ is given by Eq. 1.12.

4. Under the assumption that the ancillary qualities of a note (e.g., its attack, timbre, loudness, and duration) in an actual melody operate only to modify the weight that the note exerts in activating neurons in class $C_i$, the time-averaged activation produced by any melody $M$ in a neuron in class $C_i$ is given by Eq. 1.13.

This model may well prove to be false, in which case future experiments are likely to reject it; however, from our current vantage point, it provides a useful working hypothesis. Under the tentative assumption that the model is true, several important questions leap out:

- How many distinct neuron classes $C_i$ exist in the auditory systems of listeners with high SS?
- What do the different classes $C_i$ sense? I.e., what are the sensitivity functions $f_i(d)$ characterizing the different neuron classes $C_i$?
- How do the different ancillary properties of a note (attack, timbre, loudness, rhythmic context) operate to modify the weight exerted by the note in Eq. 1.13?

Various psychophysical methods can be used to attack these questions. By testing the relative discriminability of a sufficiently large number of tone-scrambles with different note-histograms, it is possible to determine the space spanned by the sensitivity functions $f_i$, $i = 1, 2, \cdots, N$. The dimensionality of this space is a lower bound on the number of neuron classes $C_i$ that are sensitive to scale-defined qualities. (This was essentially the method used by Maxwell (1855) to show that human color perception is 3-dimensional.) It is more challenging to determine the actual sensitivity functions $f_i$ characterizing individual classes $C_i$ of scale-sensitive neurons. A method for addressing this problem has recently been developed and applied in the domain of visual perception to analyze the mechanisms sensitive to spatially random mixtures of small squares varying in gray-scale (Silva and Chubb, 2014; Victor et al., 2017).

In order to investigate the influence of the ancillary properties of a note on the weight it exerts in activating a give scale-sensitive neuron class, $C_i$, it is first necessary to construct tone-scrambles whose note-histogram "isolates" the neuron class $C_i$. In order to construct such tone-scrambles, one must first determine the sensitivity functions $f_j$ characterizing all of the neuron classes $C_j$. Then one must derive the component $\tilde{f}_i$ of $f_i$ that is orthogonal to the space spanned by all of the other $f_j$'s (e.g., by using Gram-Schmidt orthogonalization). Then in the context of a task requiring the participant to classify tone-scrambles whose histograms differ by $\tilde{f}_i$, one can vary the ancillary properties of the notes in the stimulus to measure the influence that these changes exert on performance.

### 1.6.2 Melodic contour and the ending-on-a-high-note bias.

The contour of a melody (the rising and falling pattern of pitch in the melody) is an important aspect of melodic structure that has long been a focus of research (Justice and Bharucha, 2002; Quinn, 1999; Schmuckler, 1999, 2016). In the current context, if we define the pitch-height of a tone as 1, 4, 5, 8 or 13 depending on whether the tone is a $G_5$, $B\flat_5$, $B_5$, $D_6$ or $G_6$, then for $n = 1, 2, 4, 8$, the pitch contour of a stimulus in the $n$-task is the vector $S$ of length $n$ whose $t^{th}$ entry is the pitch-height of the $t^{th}$ tone in the stimulus. As revealed by the analysis of the 1-task in Sec. V., the pitch contours of the stimuli in all of the 1-, 2-, 4- and 8-tasks strongly influenced the judgments of most listeners. This influence obeys a simple rule. For the $n$-task, there exists a pitch-contour sensitivity profile $w = (w_1, w_2, \cdots, w_n)$ common to all listeners, and the additive bias exerted on a L's decision statistic is the inner product of $w$ with $S$.

Strikingly, this effect operates with equal strength across all listeners except those with the highest levels of SS. This suggests that sensitivity to this feature of the stimulus is independent of SS.

This observation raises the possibility that, in general, sensitivity to melodic contour is independent of SS. To investigate this issue, one would need to devise new tasks to measure sensitivity to melodic contour. For this purpose, one could use tone-scramble-like stimuli. However, to assess sensitivity to melodic contour, one would want to use the same note histogram across all trials while varying the temporal order of the notes in the stimulus. For example, one might use stimuli comprising random permutations of 13, 65ms tones, one each of the 13 notes in the chromatic scale from $G_5$ to $G_6$. In a given task condition, the listener would strive, on each trial, to judge (with feedback) whether the pitch contour of the stimulus correlated positively vs. negatively with a particular "target" contour that was fixed across all trials in the condition. This target contour might resemble the contours shown in Fig. 1.9 or it might take a different form.

Contour-sensitivity tasks of this sort can be used to probe the question of whether or not sensitivity to melodic contour is independent of SS. A natural approach is to measure performance in both the 8-task (to gauge SS) and in a contour-sensitivity task across a large sample of listeners. If the correlations between performance in the two tasks are low, this would lend support to the claim that contour-sensitivity is independent of SS.

### 1.6.3 Does musical training increase SS?

The current results echo those of Dean and Chubb (2017) in suggesting that musical training may heighten SS in some listeners but not in others (Fig. 1.5). The claim that SS can be heightened in some listeners is supported by the observation that in our sample, at least 5 years of musical training were required to attain the highest levels of SS observed.

On the other hand, the claim that musical training fails to heighten SS in other listeners is supported by the observation that our sample contained many listeners with 5 or more years of experience who nonetheless possessed low levels of SS.

The current results thus suggest the following picture: some listeners possess the potential to attain high levels of SS whereas others do not. Musical training can heighten SS only in listeners who possess this potential.

### 1.6.4 The importance of measuring SS in assessing the effects of musical training.

Trained musicians have been shown to perform better than non-musicians on a wide range of auditory tasks. For example, musicians are better than non-musicians at discriminating simple tones (Buss et al., 2014; Fujioka et al., 2004, 2005; Micheyl et al., 2006) and complex melodic stimuli (Pantev et al., 1998). They also perform better than non-musicians in tasks requiring sound segregation (Parbery-Clark et al., 2009), auditory attention (Strait et al., 2010), speech-processing (Besson et al., 2011a,b; Marie et al., 2010, 2011; Morrill et al., 2015; Parbery-Clark et al., 2011), as well as executive control (Bialystok and DePape, 2009; Zuk et al., 2014).

If it is found that musicians perform better than non-musicians in a given task, it is tempting to leap to the conclusion that musical training improves performance in the task; however, another possibility is that the task in question requires resources that are inherited or acquired through early experience, and people who possess those resources are more likely to become musicians than people who do not. It has proven challenging to decide between these alternatives in many cases.

In any study investigating the relationship between musical training and performance in some perceptual or cognitive task, we propose that the SS of all listeners should be measured as a matter of standard practice for the following reasons:

1. If performance in the target task is correlated positively with musical training, this may be due to the fact that task performance depends on SS.

2. Although SS and years of musical training are correlated, the two variables can be readily dissociated due to the existence of listeners with little or no musical training but high SS and other listeners with many years of musical training but very low SS.

3. It is easy to measure a variable that reflects SS by testing a listener in 100 trials of the 8-task and estimating $d'$ from the last 50 trials.

4. Only by including SS in one's model can one determine whether musical training accounts for any additional variance in predicting task performance.

## 1.7 Acknowledgments.

This work was supported by NIH Training Grant T32-DC010775 awarded to S. M. from the UC Irvine Center for Hearing Research.

# Chapter 2

# Can training improve performance in classifying major versus minor tone-scrambles?

## 2.1 Introduction

Several studies have provided evidence that, on average, listeners hear music in the major scale as sounding “happy” and music in the minor scale as sounding “sad” (Bonetti and Costa, 2019; Crowder, 1984, 1985a,b; Cunningham and Sterling, 1988; Gagnon and Peretz, 2003; Gerardi and Gerken, 1995; Heinlein, 1928; Hevner, 1935; Kastner and Crowder, 1990; Leaver and Halpern, 2004; Peretz et al., 1998; Temperley and Tan, 2013); however, in many of these studies, effect sizes are surprisingly modest. For example, Hevner (1935) presented listeners with 10 passages of classical music that had been recomposed into major and minor versions, where a given listener heard only one of the two versions. After hearing a given passage, the listener went through a list of adjective-pairs, selecting the word in each pair that best applied to the passage. The adjective “happy” was chosen over “sad” by 70.4% of the listeners who heard the major versions, and the adjective “sad” was chosen over “happy” by 62.6% of the listeners who heard the minor versions. Taking “happy” (“sad”) as the correct response for the major (minor) version of a given passage, the percent correct observed by Hevner (1935) is 66.5%. The other adjective pairs all yielded lower percents correct. This average is significantly higher than chance; however, it is far from perfect.

Other studies have shown that both listeners with and without musical training find it surprisingly difficult to discriminate music in the major and minor modes. Halpern (1984) asked listeners to rate the similarity/difference of pairs of melodies. Melodies that differed in rhythm were rated as being highly dissimilar; by contrast, melodies with identical rhythms that differed in mode (with one being major and the other minor) were rated as being highly similar. In a later study, Halpern et al. (1998) asked listeners to rate whether two melodies were similar or different. If the melodies differed in either rhythm or contour, listeners tended to rate them as highly different; however, if the melodies differed only in mode, listeners tended to rate them as highly similar.

Leaver and Halpern (2004) tested 43 non-musicians in several tasks investigating sensitivity to the difference between melodies in the major and minor modes. In one task, the listeners heard two melodies and had to judge whether they were the same or different (aside from being in a different key). Melodies that differed in either rhythm or contour were easily judged to be different, but melodies that differed only in mode were much more difficult to discriminate. In another task, listeners heard one melody and had to classify it based on mode. The stimuli for this task were 40 obscure folk-song melodies (20 major and 20 minor). One group of listeners was asked to classify these melodies as “major” or “minor”; another group was asked to classify them as “happy” or “sad.” The mean performance of the group tasked with classifying them as “major” or “minor” did not differ significantly from chance; strikingly, however, the group that classified them as “happy” or “sad” achieved 84%

correct (SD 7%). It should be noted, however, that this stimulus set was not constructed to isolate mode as the sole influence on classification performance. The minor (major) melodies were not recomposed into major (minor) variants; thus, it is possible that the melodies that tended to be classified as "happy" differed from those that tended to be classified as "sad" in uncontrolled ways. Indeed, the dramatic difference in performance in the "happy"/"sad" versus "major"/"minor" conditions argues that there may have been factors other than mode operating to promote the observed classifications.

Evidence is accumulating that the low average sensitivity observed in previous studies may reflect a bimodal distribution in performance, with some listeners high and others very low in sensitivity. Blechner (1977) and Crowder (1985a) (who replicated and extended Blechner's results) were the first to observe this bimodal distribution. In both studies, stimuli were triadic chords that varied in small steps from pure major to pure minor. The base and upper notes of the triad (which differed by a perfect fifth) were fixed across trials, and the middle note was varied in small steps from a minor third above the base note to a major third above the base note. The task was to judge whether the stimulus was closer to pure major or pure minor. For each listener, a psychometric function (plotting proportion correct as a function of middle-note frequency) was estimated. In each of the studies of Blechner (1977) and Crowder (1985a), listeners tended to fall into two distinct categories: for some listeners, the psychometric functions were very steep (showing that these listeners were very sensitive to the difference between the major and minor triads); for other listeners, however, the psychometric functions were close to flat (showing that these listeners could detect little if any difference between the two types of triads).

Chubb et al. (2013) found additional evidence for this bimodal distribution in sensitivity to mode variations using stimuli called "tone-scrambles," i.e., rapid, randomly ordered sequences of musical notes, all of the same loudness, timbre and duration. As emphasized by theories of music composition, the qualities that music can achieve by variations in scale are central to its meaning (Rameau, 1722; Schoenberg, 1922; Tymoczko, 2011). The original motivation for using tone-scrambles as psychophysical stimuli was to isolate such qualities from other aspects of musical structure. Although tone-scrambles differ dramatically from actual music, they provide a tool for selectively engaging the neural systems that are differentially activated by variations in musical "scale." We use the term "musical scale" very broadly to refer to the histogram of the notes the music contains. This is not to say that tone-scrambles may not differentially activate systems sensitive to other aspects of musical structure (e.g., systems sensitive to note order); however, in typical tasks that use tone-scrambles (such as the "3-task" described below), response correctness depends only on the histogram of the notes in the stimulus. Thus, to achieve optimal performance, the listener must use only those systems sensitive to variations in note histogram to make their judgments.

In the "3-task" (used in Exp. 1 of Chubb et al. (2013)), each tone-scramble contained eight copies of each of the notes $G_5$ (low tonic), $G_6$ (high tonic), $D_6$ (degree 5 of both the $G$ major and $G$ minor scales). In addition, major tone-scrambles included 8 copies of $B_5$ (degree 3 of the $G$ major scale) whereas minor tone-scrambles included 8 copies of $B\flat_5$ (degree 3 of the $G$ minor scale). Each note was a pure tone of duration 65 ms, and the 32 tones in a given stimulus were presented in random order. Thus notes occurred at the rate of 923 bps, and stimulus duration was 2.08 sec. The task was to judge (with trial-by-trial feedback) whether the stimulus was major or minor.

The results revealed a bimodal distribution in performance which has been replicated in several other studies (Dean and Chubb, 2017; Ho and Chubb, 2020; Mann, 2014; Mednicoff et al., 2018; Waz and Chubb, 2019). The histogram of proportion correct in the 3-task pooled across the 506 listeners tested in Chubb et al. (2013); Dean and Chubb (2017); Waz and Chubb (2019); Mann (2014); Mednicoff et al. (2018) is shown in Fig. 2.1. As this figure demonstrates, most listeners (approximately 70%) perform near chance; the other 30% are near perfect.

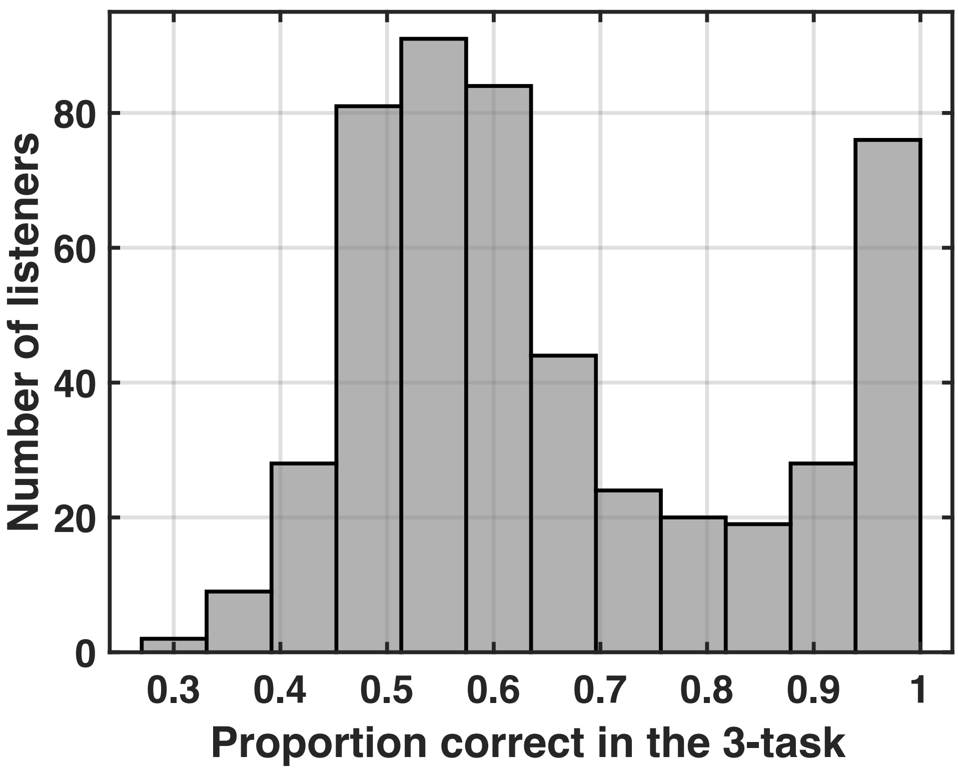


Figure 2.1: *Histogram of proportion correct in the 3-task.* Results (for 506 listeners) are pooled from Chubb et al. (2013); Dean and Chubb (2017); Waz and Chubb (2019); Mann (2014); Mednicoff et al. (2018). For each listener, proportion correct is estimated from the last 50 trials he-or-she performed. All listeners performed at least 40 trials prior to these last 50.

### 2.1.1 How do high and low 3-task performers differ in their processing capabilities?

The bimodal distribution in 3-task performance (Fig. 2.1) indicates that high-performers possess processing capabilities that low-performers lack. Dean and Chubb (2017) explored the nature of these capabilities; they tested listeners in the 3-task as well as four other similar tone-scramble tasks. All stimuli in each task contained 8 each of the notes, $G_5$, $D_6$ and $G_6$ to establish $G$ as the tonic. In addition, each stimulus contained 8 identical signal tones. In the "2-task," the signal-tone note was either $A\flat_5$ or $A_5$; in the "4-task," the signal-tone note was either $C_6$ or $C\sharp_6$; in the "6-task," the signal-tone note was either $E\flat_6$ or $E_6$; and in the "7-task," the signal-tone note was either $F_6$ or $G\flat_6$. In each case, the task was to classify the stimulus according to which of the two signal-tone notes it contained.

Dean and Chubb (2017) were able to model their results in terms of a single processing resource $R$. Specifically, their "bilinear" model proposed that the sensitivity (as reflected by $d'$) of a given listener $k$ to the difference between the two stimulus types in a given task $t$ was

$$d'_{k,t} = R_k F_t \tag{2.1}$$

where $R_k$ is the level of $R$ possessed by listener $k$, and $F_t$ is the strength with which performance in task $t$ is facilitated by $R$. The bilinear model (Eq. 2.1) accounted for 79% of the variance in $d'_{k,t}$ for 139 listeners across the 2-, 3-, 4-, 6- and 7-tasks. Dean and Chubb (2017) concluded that $R$ predominates in controlling performance across these five tone-scramble tasks.

What do these findings suggest about the nature of the resource $R$? The 3-, 6- and possibly the 7-tasks require differential sensitivity to the major vs. the minor scale. In each of these tasks, one signal-tone note belongs to the major but not the minor scale; the other belongs to the minor but not the major scale. However, in the 2- and 4-tasks, one signal-tone note belongs to both the major and minor scales, and the other belongs to neither scale. Dean and Chubb (2017) concluded that the resource $R$ is not specific to the difference between major vs. minor; rather, they proposed that $R$ confers sensitivity more generally to the spectrum of qualities that music can achieve by establishing a tonic and selecting a subset of notes to serve as a scale relative to that tonic. This led Dean and Chubb (2017) to call $R$ "scale-sensitivity," using the word "scale" very generally to refer to any probability distribution on a set of notes. They intended the term "scale-sensitivity" to refer to sensitivity to the qualities that music can achieve by varying this probability distribution. This interpretation accords with the subjective impressions of high-performing listeners who typically report that they base their judgments on qualities evoked by the tone-scrambles that mirror those evoked by music. For example, high-performers report anecdotally that in the 3-task, major and minor tone-scrambles have the "happiness" and "sadness" characteristic of music in the major (ionian) and minor (aeolian) modes.

Under this view, high and low 3-task performers are likely to differ dramatically in their experience of actual music: high-performers are assailed by variations in scale-defined qualities that low-performers are unable to hear. There are, however, reasons to doubt this interpretation.

### 2.1.2 Are high 3-task performers really sensing scale-defined qualities?

It might be objected that high-performers may not be using scale-defined qualities to make their judgments in tone-scramble tasks. In all of the tone-scramble tasks used in previous studies, the context tones (the $G_5$'s and $D_6$'s and $G_6$'s) are fixed across all trials, and all other tones in any stimulus are identical in frequency, taking a value equal to one of two possible signal notes. This raises the possibility that high-performing listeners may be ignoring the context tones and simply basing their judgments on the presence vs. absence in the stimulus of one of the two signal-tone notes on each trial.

An experiment of Waz and Chubb (2019) makes this possibility unlikely. The notes $B_5$ and $E_6$ are scale degrees 3 and 6 of the $G$ major diatonic scale whereas $B\flat_5$ is scale degree 3 of the $G$ minor scale, and $E\flat_6$ occurs in three of the four common versions of the $G$ minor scale (natural minor, harmonic minor and descending melodic minor). Thus, in the context common to all tone-scrambles in both the 3- and 6-tasks, $B_5$ and $E_6$ might both be expected to heighten stimulus "majorness" whereas $B\flat_5$ and $E\flat_6$ might be expected to heighten stimulus "minorness". Dean and Chubb (2017) found that $F_{\text{3-task}} \approx F_{\text{6-task}}$ in

Eq. 2.1 implying that performance in the 3- and 6-tasks was nearly equal. Thus, if high-performing listeners are basing their judgments on the "majorness" vs. "minorness" of 3- and 6-task stimuli, then major stimuli should differ from minor stimuli along this spectrum by roughly the same amount in the 3- and 6-tasks.

In the experiment of Waz and Chubb (2019), each of the 100 listeners was tested in seven tone-scramble tasks. All stimuli in all tasks comprised of the same 24 context tones: 8 $G_5$'s, 8 $D_6$'s and 8 $G_6$'s, establishing $G$ as the tonic. For our purposes, it is useful to focus on four of the seven tasks. The first two of these are the 3- and 6-tasks (Dean and Chubb, 2017). As a reminder, in the 3-task, the eight signal tones in a given stimulus are either all $B_5$'s or else all $B\flat_5$'s; and in the 6-task, the eight signal tones in a given stimulus are either all $E_6$'s or else all $E\flat_6$'s. The other two tasks are called "hybrid" tasks because stimuli include mixtures of different signal-tone notes. In the 3u6-task (the "uncrossed 3,6 task"), in addition to the 24 context tones, every Type-1 tone-scramble contained 4 $B_5$'s and 4 $E_6$'s, and every Type-2 stimulus contained 4 $B\flat_5$'s and 4 $E\flat_6$'s. In the 3x6-task (the "crossed 3,6 task"), in addition to the context tones, Type 1 stimuli contained 4 $B_5$'s and 4 $E\flat_6$'s whereas Type 2 stimuli contained 4 $B\flat_5$'s and 4 $E_6$'s.

If high-performing listeners ignore the context tones and base their judgments on the presence vs. absence of tones of a particular signal frequency in the stimulus, then both of the 3u6- and 3x6-tasks should be harder than either of the 3- or 6-tasks because any specific signal frequency that the listener might choose to listen for occurs only half as many times in hybrid stimuli vs. non-hybrid stimuli. On the other hand, if high-performing listeners base their judgments on the majorness vs. minorness of the stimulus, then

1. the 3u6-task should yield performance comparable to the 3- and 6-tasks because in this task (as in the 3- and 6-tasks) Type 1 (Type 2) stimuli contain eight signal tones all of which operate with roughly equal effectiveness to heighten "majorness" ("minorness").

2. the 3x6-task should yield performance that is suppressed in comparison to the 3- and 6-tasks because in this task each of Type 1 and Type 2 stimuli contain four major and four minor signal tones. Thus, these stimuli should produce qualities that are roughly equated in their levels of majorness vs. minorness, forcing high-performing listeners to base their judgments on one or another secondary, scale-defined quality that is unlikely to support the same level of performance as we see in the 3- and 6-tasks.

In summary, if high-performing listeners base their judgments on the presence vs. absence of tones of a particular frequency, performance should be suppressed in both of the 3u6- and 3x6-tasks compared to the 3- and 6-tasks. By contrast, if high-performing listeners base their judgments on stimulus majorness vs. minorness, then performance in the 3x6-task should be suppressed in comparison to the other 3 tasks, which should all afford roughly equal performance.

The findings support the latter prediction. The results of Waz and Chubb (2019) are well-described by the bilinear model (Eq. 2.1) which accounts for 81% of the variance in $d_{k,t}$ for 100 listeners across 7 tone-scramble tasks. This implies that a single processing resource predominates in controlling performance across all of these tasks.

For current purposes, the important result is that $F_{\text{3-task}} = 1.15$, $F_{\text{6-task}} = 1.11$, and $F_{\text{3u6-task}} = 1.14$, whereas $F_{\text{3x6-task}} = 0.67$. (For all four tasks $t$, 95% Bayesian credible intervals are approximately $F_t \pm 0.04$.) Thus, performance in the 3x6-task is suppressed by roughly 41% in comparison to the 3-, 6- and 3u6-tasks, all of which yield performance that is equal within measurement error. We conclude that high-performing listeners are in fact using scale-defined qualities to make their judgments in the 3-, 6-, 3u6- and 3x6-tasks.

### 2.1.3 What is the source of the bimodal distribution in 3-task performance?

Several possible explanations for the bimodal distribution shown in Fig. 2.1 have been ruled out. First, the tones in a tone-scramble are very rapid (923 bpm); one might therefore wonder whether high-performers in the 3-task differ from low-performers merely in being able to extract modal qualities from such rapid stimuli. The answer is no. Mednicoff et al. (2018) tested listeners in the basic 3-task described above as well as three other, slower 3-task variants. In the slowest variant, each tone lasted 520 ms, and each stimulus contained only four, randomly ordered tones, one each of the notes $G_5$, $D_6$, $G_6$ and either a single $B_5$ (in major stimuli) or a single $B\flat_5$ (in minor stimuli). This variant of the 3-task proved to be the hardest of the four, significantly more difficult than the three faster variants (which afforded roughly equal performance). Mednicoff et al. (2018) concluded that it is not the high rate of presentation in the 3-task that blocks low-performers from sensing the difference between major and minor tone-scrambles.

Second, the bimodal distribution shown in Fig. 2.1 is not produced by musical training. Fig. 2.2 plots Years-of-musical-training vs. 3-task-$d'$ for the same 504 listeners whose performance histogram is shown in Fig. 2.1. Although performance in the 3-task is positively correlated with years of musical training ($r = 0.44$, $p = 0.0000$), this correlation is driven mainly by a large group of listeners with no musical training whose 3-task-$d'$ values are near 0. However, there also exist (1) some listeners with many years of musical training for whom 3-task-$d' \approx 0$, showing that musical training does not suffice to produce high performance, and (2) other listeners with little or no musical training who achieve 3-task-$d'$ values near ceiling showing that musical training is not necessary for high performance.

Indeed, Fig. 2.2 is consistent with the idea that skill in the 3-task is largely immune to musical training. Under this story, sensitivity in the 3-task is fixed early in life, and the positive correlation between musical training and 3-task-$d'$ is due to the fact that listeners with high sensitivity are more likely to seek out musical training than listeners with low sensitivity. This idea receives additional support from the finding that 6-month-old infants show the same bimodal distribution of performance in the 3-task as adults (Adler et al., 2020).

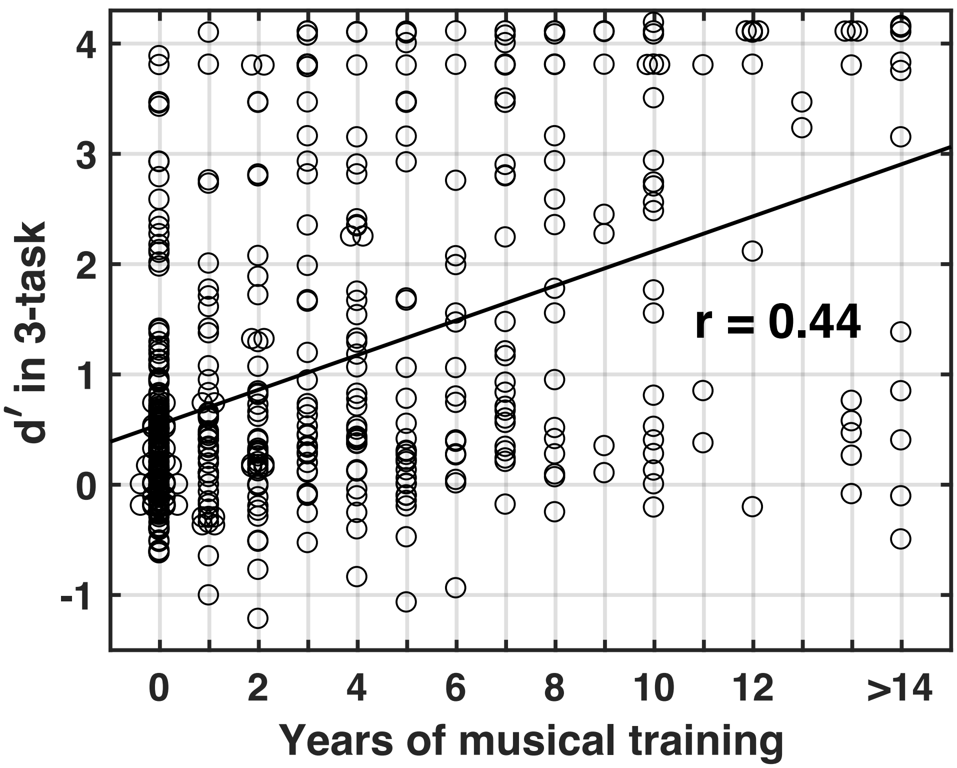


Figure 2.2: *Scatter plot of years-of-musical-training vs. 3-task-$d'$.* This figure pools results from Chubb et al. (2013); Dean and Chubb (2017); Waz and Chubb (2019); Mann (2014); Mednicoff et al. (2018). There are 506 listeners included in this sample. For each listener, proportion correct is estimated from either 45 or 50 trials.

On the other hand, there is evidence that training can improve 3-task performance for at least some listeners. First in Fig. 2.2, there are many listeners with 0 years of musical training, and none of them achieve perfect performance ($d' = 4.1075$) in the 3-task. By contrast, many of the listeners with one or more years of musical training perform perfectly, suggesting that musical training can elevate 3-task performance at least slightly in some listeners. Other evidence comes from Chubb et al. (2013). Specifically, in Experiment 2 of that study, 104 listeners were tested in four 50-trial blocks of a variant of the 3-task. (This task differed from the standard 3-task in that the 16 $G$'s in each stimulus comprised randomly varying numbers of $G_5$'s and $G_6$'s.) Thirty-eight of the listeners tested in this task achieved proportions correct between 0.6 and 0.9. For these listeners, $d'$ increased significantly on average across the four blocks (with the largest increase occurring on block 4).

However, it should be noted that 42 listeners in this experiment achieved average proportions correct less than 0.6; therefore, these listeners did not exhibit learning across the four blocks. It is possible that for some or all of these listeners, no training regimen exists that can improve their 3-task performance. It is also possible, however, that these listeners do possess neural resources sufficient to achieve high levels of 3-task performance which they have thus far failed to access. Although these low-performing listeners are not immediately assailed by an obvious, qualitative difference between major and minor 3-task stimuli, they may be able to tune in to, and perhaps amplify, this difference if their attention is directed to the difference between stimuli. The current experiment investigates this possibility.

### 2.1.4 The current study

In all of the previous studies using the 3-task, (Chubb et al., 2013; Dean and Chubb, 2017; Waz and Chubb, 2019; Mann, 2014; Mednicoff et al., 2018) listeners experienced no more than 200 trials in total. The current study investigated whether low-performing listeners might benefit from a relatively brief (2-hour) training regimen designed to improve performance in the 3-task. In particular, if it is found that such a training regimen can boost performance in all listeners to near ceiling, this would suggest that low-performing listeners need only overcome some relatively minor processing difficulty in order to hear the same qualitative differences between major and minor tone-scrambles as high-performing listeners. Such a finding would suggest that the bimodal distribution in performance (Fig. 2.1) does not reflect an important difference in perceptual potential between high- vs. low-performing listeners. On the other hand, if a large proportion of low-performing listeners fails to benefit from such a training regimen, this will bolster the claim that the processing potential of these listeners differs importantly from that of high-performers.

## 2.2 Methods

All methods were approved by the UCI Institutional Review Board.

### 2.2.1 Listeners.

Seventy-one undergraduates from the University of California at Irvine participated. All had self-reported normal hearing. Listeners were compensated with course credit. Forty-three had one or more years of formal musical training (mean 3.83 years). An additional 44 listeners were excluded from the analysis because they either failed to return for day-2 of the experiment ($n = 28$) or because their performance after 200 training trials in the regimen described below suggested that they were unlikely to benefit from the training regimen ($n = 16$). The criterion used to window out the latter class of listeners is described below.

### 2.2.2 Stimuli.

The stimuli were tone-scrambles. Each tone was a 65ms pure tone windowed by a raised cosine with at 22.5 ms rise and fall time. All stimuli were constructed from the notes $G_5$ (783.99 Hz), $D_6$ (1174.66 Hz), $G_6$ (1567.98 Hz), $B_5$ (987.77 Hz) and $B\flat_5$ (932.33 Hz) from the standard, equal-tempered scale. Any given stimulus contained multiple copies of the notes $G_5$, $D_6$ and $G_6$; these notes served to establish $G$ as the tonic of each stimulus. In addition, major stimuli contained one or more copies of the note $B_5$ (and no occurrences of $B\flat_5$) whereas minor stimuli contained one or more copies of the note $B\flat_5$ (and no occurrences of $B_5$). The particular numbers of these different notes that occurred in a given stimulus were varied over the course of training as described below, but each stimulus contained a total of 32 tones. In any given trial throughout the experiment, the order of the tones in the stimulus was randomized.

### 2.2.3 The Basic task.

On any given trial, the listener heard a single tone-scramble and strove to classify it as major vs. minor. The listener entered a "1" on the keyboard if he-or-she judged the stimulus to be major or a "2" if minor. Correctness feedback was given visually on the computer monitor after every trial. To take advantage of the finding of Leaver and Halpern (2004) that non-musicians are better able to classify melodies as "happy" vs. "sad" than they were to classify melodies as "major" vs. "minor," we used a response prompt that encouraged listeners to think of major tone-scrambles as "happy" and minor tone-scrambles as "sad". Specifically, participants read the prompt "Enter 1 for Major (HAPPY) or 2 for Minor (SAD)."

### 2.2.4 Procedure.

Listeners participated for one hour on each of two consecutive days. The experimental protocol is outlined in Table 3.1. On day 1, the listener first performed a block of 50 practice trials, then a 50-trial pretest, then six 100-trial training blocks, and a final 50-trial posttest. On day 2, the listener performed six more 100-trial training blocks and a final 50-trial posttest.

Table 2.1: *Procedure overview.*

| | **Day One** | **Day Two** |
|---|---|---|
| **Practice** | 1 Block of 50 Trials | – |
| **Pretest** | 1 Block of 50 Trials | – |
| **Training** | 6 Blocks of 100 Trials | 6 Blocks of 100 Trials |
| **Posttest** | 1 Block of 50 Trials | 1 Block of 50 Trials |

#### 2.2.4.1 The outset.

Each listener completed a questionnaire at the beginning of the experiment. The only information that was used from this questionnaire was the number of years of musical training of each listener. Stimuli were presented through JBL Elite 300 noise-cancelling headphones with volume adjusted to comfort level prior to testing. Testing was done in a quiet lab space on a Windows Dell computer with a standard Realtek audio/sound card using Matlab. Before the initial 50 practice trials, listeners heard eight, visually-identified example stimuli that altered between the major and minor tone-scrambles.

#### 2.2.4.2 The practice block.

On each trial in the initial practice block, the tone-scramble stimulus contained eight each of the notes $G_5$, $D_6$ and $G_6$. In addition, each major tone-scramble contained eight $B_5$'s whereas each minor tone-scramble contained eight $B\flat_5$'s. These stimuli are identical to those used in Exp. 1 of Chubb et al. (2013); they have also been used in Dean and Chubb (2017) and Mednicoff et al. (2018).

#### 2.2.4.3 The pretest and posttests.

The fifty trials in the pretest and in each of the Day-1 and Day-2 posttests included 25 major and 25 minor tone-scrambles. These were independently generated in the pretest and in each posttest. We used a more challenging task in the pretest and in each of the Day-1 and Day-2 posttests than had been used in previous studies (Chubb et al., 2013; Dean and Chubb, 2017; Mednicoff et al., 2018). On each trial in the pretest and each posttest, the tone-scramble stimulus contained 10 each of the notes $G_5$ and $G_6$ and 11 copies of $D_6$. In addition, each major tone-scramble contained a single $B_5$ whereas each minor tone-scramble contained a single $B\flat_5$. We expect many listeners to perform poorly in the pretest due to the difficulty of this "single-signal-tone" version of the 3-task. We know, however, that some high-performers perform perfectly in this task.

Thus, if the training regimen proves effective, then we expect to see at least some listeners perform near perfectly in the Day-2 post-test. By using the single-signal-tone task instead of the easier 8-signal-tone task, we sought to spread out the distribution of performance after training and thereby obtain deeper insight into the distribution of post-training sensitivity.

#### 2.2.4.4 The training blocks.

As suggested by Fig. 2.1, prior to training, many listeners ($\approx$ 70%) hear little (if any) difference between major vs. minor tone-scrambles. We attempted to provide such listeners with traction in the task by introducing an ancillary cue intended to enable all listeners to discriminate major vs. minor tone-scrambles. Over the course of training, as performance improved, this cue was gradually reduced in strength so that performance would ultimately need to rely only on the major vs. minor difference.

At the outset of training, all stimuli contained eight each of the notes $G_5$, $D_6$ and $G_6$.

In addition, all major tone-scrambles contained eight $B_5$'s, and all minor tone-scrambles contained eight $B\flat_5$'s. The ancillary cue was introduced as follows: if $A$ is the amplitude of notes $G_5$, $D_6$, $G_6$ and $B\flat_5$ in all stimuli, then the starting amplitude of all $B_5$'s was $A_{max} = 10 \times A$. Thus at the outset of training, all instances of $B_5$ were 10 dB louder than the other tones. This imparted a "raggedness" to major tone-scrambles that minor tone-scrambles lacked.

The strength of this loudness cue was controlled by a 3-down-1-up staircase. In order for the $B_5$-amplitude to decrease from $A_{max}$, the listener had to respond correctly to three stimuli in a row. Thereafter, following every trial, if the previous three responses were all correct, the $B_5$-amplitude was decreased by $\frac{A}{5}$; otherwise, the $B_5$-amplitude was increased by $\frac{A}{5}$ (but never above $A_{max}$). Thus, the $B_5$-amplitude staircase included 50 linear steps separating the $A_{max}$ from $A$.

#### 2.2.4.5 The criterion for exclusion after 200 training trials.

After 200 trials of the Day-1 training session (two 100-trial blocks), if the threshold for a given listener (estimated by fitting a Weibull function to the data from the first 200 trials) was at staircase-level 40 or higher, the listener was excused from the experiment. The total number of listeners in this group was 16. The "raggedness" imparted to major stimuli (by making the $B_5$'s 10 dB louder all of the other tones in either the major or minor stimuli) seemed to us to be sufficiently obvious to enable all listeners to discriminate major from minor stimuli at the outset. This suggests that at least some of the listeners in this group were making insufficient effort. However, we cannot rule out the possibility that some of these listeners were trying hard but (despite the "raggedness" cue) were unable to hear any difference between the major and minor stimuli. Regardless of why these listeners failed to improve, it seems unlikely that any of them would have shown any benefit from the training regimen. Thus, in excusing them, we were refining our pool to include only listeners who were likely to benefit from the training regimen.

#### 2.2.4.6 Extending the staircase levels below zero.

Many listeners performed well enough in the staircase-controlled training sessions that their $B_5$-amplitudes descended all the way to $A$. In order to test the sensitivity of these listeners as thoroughly as possible, seven additional steps were included in the staircase. Across all of these additional staircase steps, all tones in all stimuli had amplitude $A$; however, the number of "signal tones" ($B_5$'s or $B\flat_5$'s) was decreased below 8. It is convenient to number the staircase steps from 50 down to -7. The stimuli corresponding to these steps obey the following rules:

1. for $k = 50, 49, \cdots, 0$,
   (a) all stimuli contain eight each of the notes $G_5$, $D_6$, $G_6$ with amplitude $A$;
   (b) all minor stimuli contain eight $B\flat_5$'s with amplitude $A$;
   (c) all major stimuli contain eight $B_5$'s with amplitude $A + \frac{kA}{5}$.
2. for $k = -1, -2, \cdots, -7$,
   (a) all 32 tones in any stimulus have amplitude $A$;
   (b) the number of $B_5$'s ($B\flat_5$'s) in all major (minor) stimuli is $8 - |k|$;
   (c) the numbers of $G_5$'s, $D_6$'s and $G_6$'s are selected randomly under the constraint that any two of them differ by at most one.

At the start of the Day-2 training session, the staircase was reset to level 50 so that the loudness cue was once more at full strength and the training procedure was repeated.

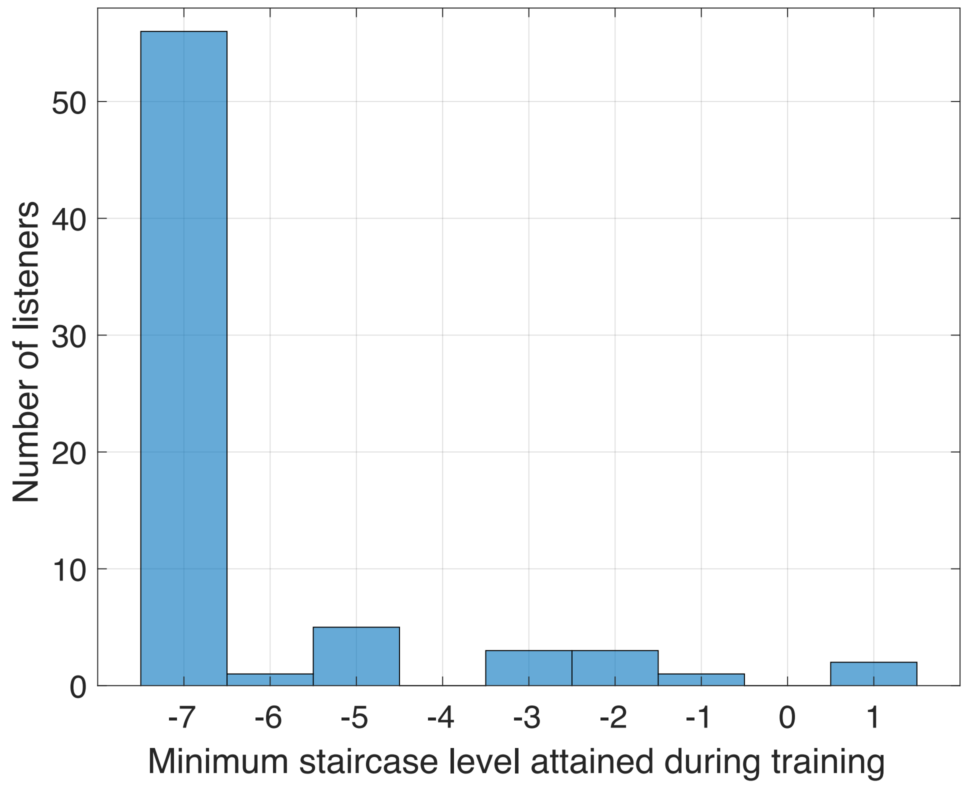


Figure 2.3: *Histogram of the minimum staircase level attained during training.* The fact that all listeners attained minimum staircase levels near or below 0 suggests that the loudness cue operated as intended.

## 2.3 Results

### 2.3.1 Training session results.

To assess the effectiveness of our training regimen, we will focus mainly on the difference in performance from the pretest to the Day-1 and Day-2 posttests. First, however, we need to check the results from the training regimen to verify that the loudness cue worked as intended. We will also look for evidence that performance improved from the Day-1 to the Day-2 training session.

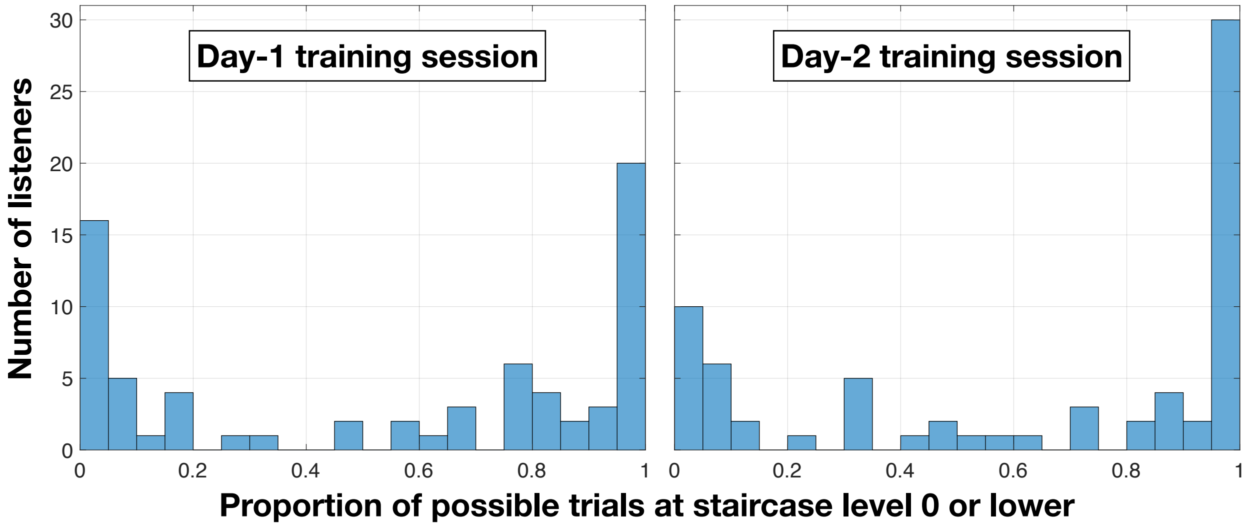


Figure 2.4: *Histogram of the proportion of possible trials at staircase level 0 or lower.* The staircase of a given listener $k$ could spend at most 547 trials in a given training session at level 0 or lower. Let $p_k$ be the proportion of those trials for which the staircase of listener $k$ was at level 0 or lower in a given training session. The left (right) panel shows the histogram of $p_k$ across all listeners $k$ for the Day-1 (Day-2) training session. Note that on Day-2 there were 10 more listeners than there were on Day-1 whose staircases remained at level 0 or below across nearly all possible trials; similarly, there were 6 fewer listeners on Day-2 whose staircases were at level 0 or lower on fewer than 5% of all possible trials.

#### 2.3.1.1 Did the loudness cue work?

If the loudness cue worked as intended, then each listener should have reached staircase levels near 0 at some point during the 1200 training trials that they performed in. The histogram of the minimum staircase levels attained by our listeners across the course of the 12 training blocks is shown in Fig. 2.3. As this figure shows, most listeners managed to reach staircase levels below 0. Thus, nearly all listeners experienced at least some training trials in which all tones in both major and minor stimuli were equal in amplitude.

#### 2.3.1.2 Listeners improved from the Day-1 to the Day-2 training session.

Each listener was trained in 600 trials during each of the Day-1 and Day-2 training sessions. On each day, at least the first 53 of those trials were at staircase levels greater than 0. Thus, the staircase of a given listener could spend at most 547 trials in a given training session at level 0 or lower. The left (right) panel of Fig. 2.4 plots the histogram (across all 71 listeners) of the proportion of the 547 trials in the Day-1 (Day-2) training session on which the listener's staircase was at level 0 or lower. There are obvious signs that performance improved from Day 1 to Day 2. In particular, on Day 2, there were 10 more of the highest-performing listeners (i.e., listeners whose staircases remained at level 0 or below on over 95% of all 547 possible trials); there were also 6 fewer of the lowest-performing listeners on Day 2 (i.e., listeners whose staircases were above level 0 on more than 95% of all 547 possible trials).

### 2.3.2 Comparing performance across pretest and both posttests.

We measured $d'$ from the pretest and each of the Day-1 and Day-2 posttests. If a listener was perfect in any of these tests (i.e. correctly responded "major" ("minor") to all 25 major (minor) stimuli), then the probability of the correct response was adjusted to $\frac{24.5}{25} = 0.98$ (Macmillan and Kaplan, 1985). Thus, perfect performance resulted in a $d'$ value of 4.1075.

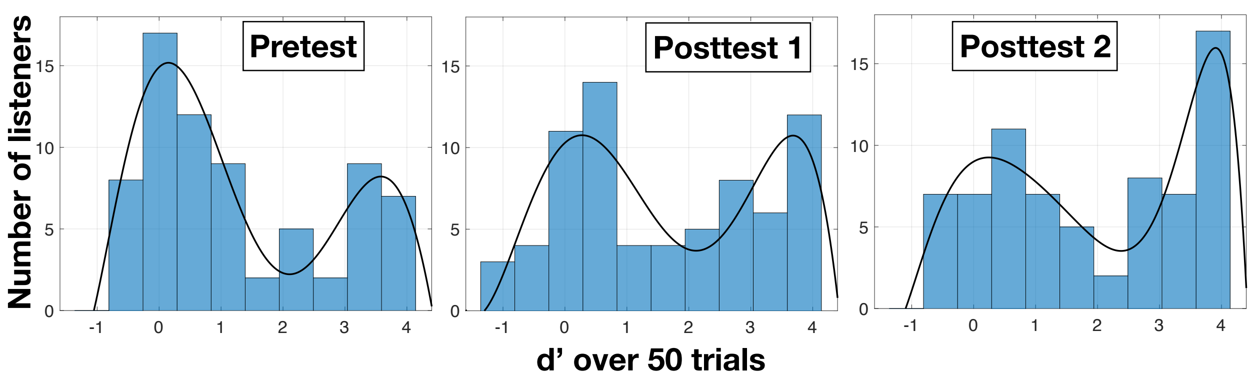


Figure 2.5: *Histograms of the distribution of performance across assessments.* $d'$ estimates are based on the 50 trials in each assessment. Note first that all three histograms are bimodal with one peak near $d' = 0$ and another near $d' = 4$ suggesting that some listeners (those with $d'$ near zero on posttest 2) fail to improve with training. Note also, however, that the peak near 0 gets smaller, and the peak near 4 gets larger across the three assessments suggesting that the performance of some listeners improves with training.

Figure 2.5 shows the histogram of $d'$ values achieved by our listeners on each of the pretest and the Day-1 and Day-2 posttests. Paired samples $t$-tests confirm that on average $d'$ values increase from the pretest to the Day-1 posttest ($t = 3.04$, df $= 70$, $p = 0.003$) and from the Day-1 posttest to the Day-2 posttest ($t = 5.186$, df $= 70$, $p < 0.001$).

Although we see clear evidence that the performance of at least some listeners improved with training, the histogram from the Day-2 posttest suggests that many listeners failed to improve. In particular, the $d'$ histogram from the Day-2 posttest remains bimodal with many listeners achieving $d'$ values near 0.

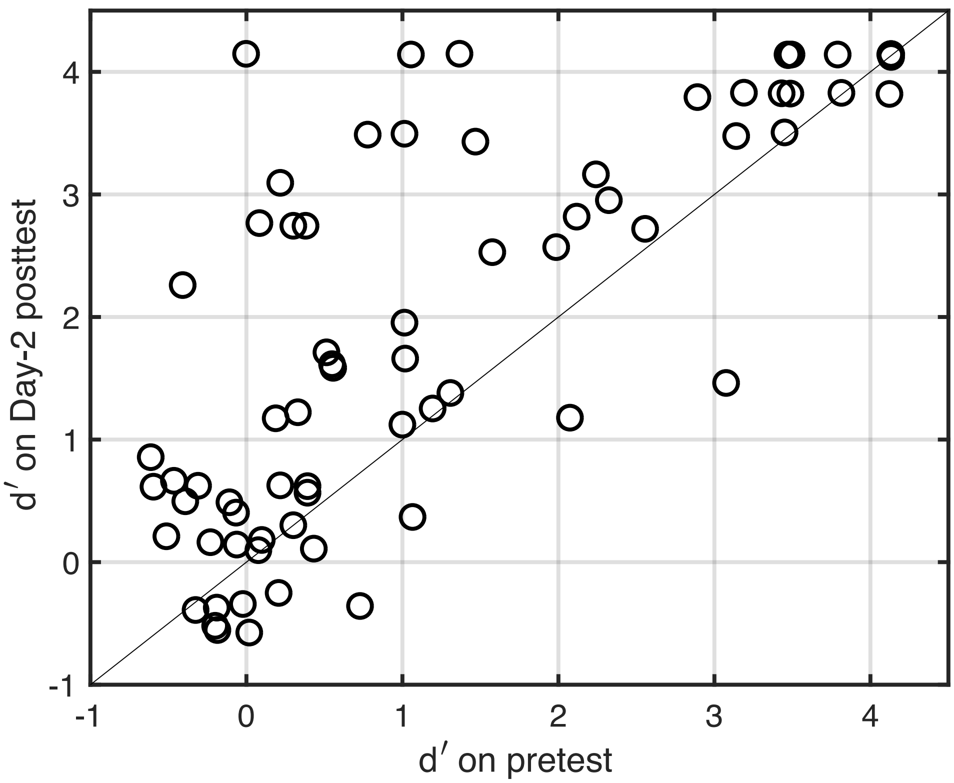


Figure 2.6: *Scatterplot comparing performance between pretest and Day-2 posttest.* Note that most listeners showed some improvement, and some listeners improved dramatically. However, many listeners who performed poorly on the pretest also performed poorly on the posttest suggesting that the training regimen did not help them.

Fig. 2.6 shows a scatterplot of $d'$ on the pretest vs. $d'$ on the Day-2 posttest. Here we see that some initially low-performing listeners improved dramatically. Note, for example, that one listener who had $d' = 0$ on the pretest achieved $d' > 4$ on the Day-2 posttest. The points corresponding to a number of other listeners are well above the main diagonal indicating that they improved substantially over the course of training. On the other hand, however, many listeners who achieved $d' < 1$ on the pretest also had $d' < 1$ on the Day-2 posttest suggesting that although training can help some listeners, for most initially poor-performing listeners, this is not true.

This point is dramatized by Fig. 2.7 which plots a histogram of the improvement in $d'$ from pretest to the Day-2 posttest for all listeners who had $d' < 1$ on the pretest. Note that 30 out of the 37 listeners who performed poorly on the pretest show little evidence of improvement due to training.

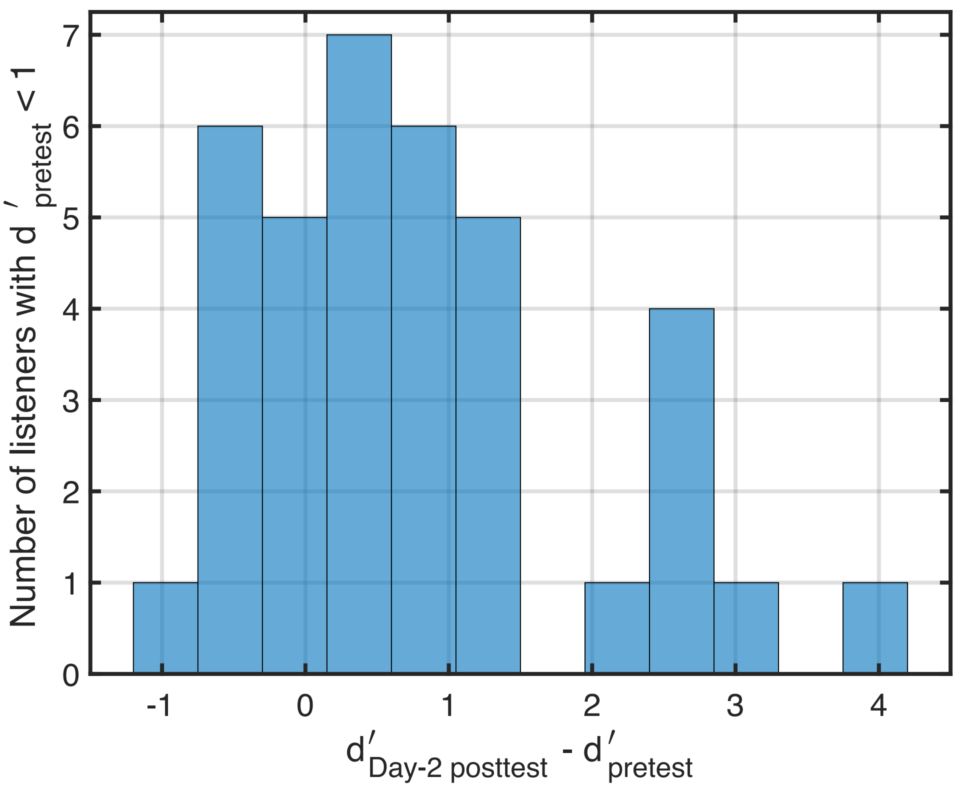


Figure 2.7: Difference in performance between the pretest and the Day-2 posttest for listeners with pretest $d' < 1$.

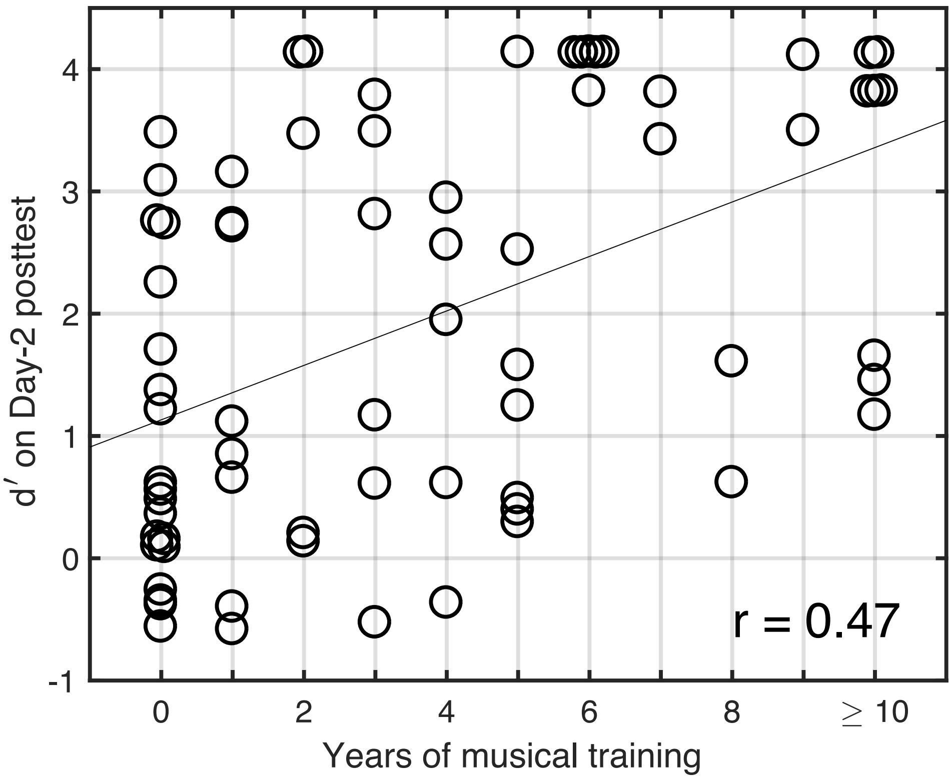


Figure 2.8: *Scatterplot of performance vs. years of musical training.*

### 2.3.3 The relationship between musical training and performance

Fig. 2.8 shows a scatterplot relating years of musical training to $d'$ on the Day-2 posttest. The diagonal line is the regression line, and the correlation of 0.47 is significantly greater than 0 ($p < 0.0001$). It should be noted, however, that this correlation is driven primarily by the large number of listeners with no musical training who perform poorly on the Day-2 posttest. Our sample also contains a number of listeners with little or no musical training who perform well on the Day-2 posttest as well as other listeners with substantial musical training who perform relatively poorly on the Day-2 posttest. Thus (consonant with findings in previous studies), musical training is neither necessary nor sufficient to insure that the training regimen we have used in the current study can elevate performance in the tone-scramble task to a high level.

## 2.4 Discussion

### 2.4.1 Most low performers do not benefit from the training regimen tested here.

Previous studies have shown that most listeners ($\approx 70\%$) perform near chance in classifying major vs. minor tone-scrambles (Chubb et al., 2013; Dean and Chubb, 2017; Mann, 2014; Mednicoff et al., 2018; Waz and Chubb, 2019; Ho and Chubb, 2020). The inability of low-performing listeners to perform the 3-task (or any of the other tone-scramble tasks used in Dean and Chubb (2017) or Waz and Chubb (2019)) above chance implies that they are unable to extract scale-defined qualities from tone-scrambles. However, Waz and Chubb (2019) have presented evidence that high-performing listeners do indeed base their responses in the 3-, 6-, 3u6- and 3x6-tasks on scale-defined qualities (as opposed to listening for the presence vs. absence of a specific signal frequency). Thus high- and low-performers experience tone-scrambles very differently; for high-performers, tone-scrambles evoke a range of scale-defined qualities that are not available to low-performers. This raises the possibility that low-performers may be afflicted by a general deficit in sensitivity to scale-defined qualities, a deficit that affects not only their experience of tone-scrambles but also of actual music.

Such a contention would be undermined if it could be shown that a modest training intervention worked effectively to eliminate the difference between low- vs. high-performing listeners. The current results show that for most low-performing listeners, the particular training regimen we have tested does not alter performance.

This is seen most clearly in Fig. 2.7 which looks specifically at the change in $d'$ in the single-signal-tone 3-task for low-performing listeners before and after the two-hour training regimen. Although seven of these 37 low-performers show substantial improvement, the other thirty do not.

It is possible that a different training regimen might be more effective. For example, if the raggedness cue more naturally aggregates with minorness than with majorness, then the training regimen might be more effective if the raggedness cue is imparted to the minor (instead of the major) stimuli. However, we suspect that such regimen-dependent variations in effectiveness will be small. In particular, the correctness feedback that was provided after every trial throughout all phases of practice, testing and training should insure that performance is limited primarily by the processing resources available to the listener rather than by details of stimulus design. We doubt, therefore, that any two-hour training regimen will be much more effective than the one tested here. However, in many psychoacoustic studies, substantial training is required to achieve asymptotic performance levels (Leek and Watson, 1984, 1988); thus, whether a long-term training regimen might be able to raise the performance of some low-performing listeners to ceiling remains an open question.

### 2.4.2 Why do some low-performers improve while others do not?

A brief training regimen such as the one tested here is unlikely to significantly alter the functionality of neural systems used to respond to differences in auditory qualities. We therefore assume that any listener who performed near ceiling on the Day-2 posttest had access at the time of the pretest to the same, highly effective neural system that they used in the Day-2 posttest.

Our working hypothesis, therefore, is that for listeners in this class, the entire effect of the training regimen is to clarify which properties to attend to when making their judgments. Under this story, all listeners who perform near ceiling on the Day-2 posttest are endowed with equally effective computational resources that are invariant across training; however, some of these listeners perform poorly prior to training because they use the wrong resources to determine their responses.

### 2.4.3 Do low 3-task performers experience real music differently from high-performers?

Evidence suggests that high 3-task performers experience a range of scale-defined qualities in tone-scrambles that low-performers cannot hear. Does this difference in sensitivity also extend to actual music? If so, then high-performers may experience a wide range of scale-defined qualities in music that low-performers cannot hear.

It should be noted, However, that tone-scrambles in the 3-task (and the other tasks used in Dean and Chubb (2017); Mednicoff et al. (2018); Ho and Chubb (2020); Waz and Chubb (2019)) lack ecological validity. They are faster than actual music (923 BPM). In contrast to real music in which notes typically have complex and variable attacks and timbres, tone-scrambles are composed of pure tones. They are also higher in pitch than most music. Perhaps this combination of non-musical features blocks low-performers from sensing any difference between major and minor stimuli in the 3-task. Although Mednicoff et al. (2018) have shown that slowing down 3-task stimuli does not suffice to enable low-performers to perform well, it is possible that low-performers may improve if the stimuli are (1) slowed down, and (2) composed of notes with (a) attack and timbre characteristic of an actual musical instrument and (b) pitches from the middle of the piano keyboard. If so, then low 3-task performers may be able to hear the same scale-defined qualities in actual music that high-performers can. Experiments are under way to test these possibilities.

## 2.5 Acknowledgments.

This work was supported by NIH Training Grant T32-DC010775 awarded to S. M. from the UC Irvine Center for Hearing Research. We would like to acknowledge Stephanie (Mejia) Hertel, Karen Gonzalez, and Juan A. Chavez for their contributions in assisting with data

collection.

# Chapter 3

# The Neurobiological Basis of Major/Minor Musical Mode Processing

## 3.1 Introduction

How and why are humans sensitive to music? Music has been found to evoke physical responses that range from rewarding "chills" (Salimpoor et al., 2011) to enhancing therapeutic treatment in Alzheimer's patients (Fang et al., 2017). Pitch is a component in both music and language that conveys information (Juslin and Laukka, 2003; Plack and Oxenham, 2005; Swaminathan and Schellenberg, 2015) which may contribute to these responses, and a growing body of research suggests that musical training may enhance linguistic abilities (Patel and Iversen, 2007; Koelsch and Siebel, 2005; Slevc, 2012), and more specifically, sensory encoding of pitch patterns as related to speech intelligibility (Patel, 2011; Strait et al., 2012) and intonation (Magne et al., 2006; Schön et al., 2004). The similarities of pitch in both music and language suggests a common processing mechanism between domains and may explain why people associate affective connotations of major and minor scales of music as “happy” and “sad,” respectively; however, a majority of the literature is in conflict about the neurobiological basis of major-minor musical processing, and whether musicians and non-musicians process music similarly or in different ways (Crowder, 1984, 1985a,b; Halpern et al., 1998; Leaver and Halpern, 2004). As noted by Virtala and Tervaniemi (2017), research that controls for systematic variations in musical mode using both chords and tone sequences and that takes careful consideration in comparing performance between musicians and non-musicians is largely absent from the literature.

The existing literature investigating the neurobiological basis of major-minor musical mode processing has found behavioral differences in musical ratings between musicians and non-musicians, yet provides no evidence for differences in underlying neural activity as a function of musical expertise. In one study, Pallesen et al. (2006) asked participants to rate each chord on two scales: unpleasant-pleasant and sad-happy. Behaviorally, they found that musicians rated minor chords as sadder and dissonant chords as more unpleasant, yet no significant differences in neural activations were seen between the two groups during passive listening. Because of these results, the authors attribute the difference to a musician’s expertise and familiarity at recognizing the mode of the stimulus. Consequently, using a univariate, general linear model typical for analyzing blood oxygen level-dependent (BOLD) signals from fMRI may be limited. Univariate analyses summarize task-related changes in voxels and regions of interest; therefore, restricting its ability to capture the full pattern of activity that may be utilized by a trained musician to discriminate between the two musical modes.

Thus, the differences in neural activity between listeners may only be revealed through multivariate pattern analyses (MVPA). Two studies, Lee et al. (2011) and Klein and Zatorre (2015), have used MVPA in music processing research. Lee et al. (2011) showed that the processing of ascending and descending musical contours was reliably categorized in three significant regions (based on searchlight analysis): the right superior temporal sulcus (STS), the intraparietal lobe (or more specifically intraparietal sulcus), and the anterior cingulate cortex (ACC). Consistent with those findings, the study conducted by Klein and Zatorre (2015) revealed significant information in the pattern of neural activations in regions of the superior temporal and intraparietal sulci that discriminated major and minor stimuli; however, only six trained musicians, who were deemed experts at the behavioral task, ultimately completed the protocol and were scanned in this study. Consequently, it is difficult to conclude based on these results whether underlying neural differences exist between musicians and non-musicians, so the open question remains: do trained musicians process musical modes in fundamentally different ways from untrained listeners?

Research that supports listeners behaviorally experiencing music in different ways has found that a majority of all listeners (about 70%) cannot discriminate between major and minor musical modes, while the other 30% performed near perfect (Chubb et al., 2013). By using a stimulus called a tone-scramble, they identified those who could and could not tell the difference between the musical modes. Surprisingly, only a weak correlation was found between performance and musical training. This weak relationship between performance in the task and musical training suggests that some listeners may possess a cognitive resource, R, due to the divide in performance across participants. Dean and Chubb (2017) investigated the nature of this cognitive resource through five other variants of the tone-scramble task using variations of semitones within the diatonic scale. Their results were well described in a proposed bilinear model, where performance in all five of the tasks was facilitated by a listeners' cognitive resource, R. The model accounted for 79% of the variance in performance across 139 listeners in all five tasks. Again, most listeners (around 70%) had R values around 0, corresponding to near-chance performance, while 30% of listeners yield much higher levels of R near perfect. From these results, Dean and Chubb (2017) concluded that this resource can be described as "scale-sensitivity." Therefore, another open question that remains is: do listeners high in scale-sensitivity show differences in neural activation from listeners low in scale-sensitivity?

The current study aims to fill this gap by testing the extent to which listeners differ in musical sensitivity regardless of their musical training. This study builds upon the previous behavioral data to investigate the neurobiological basis of major-minor musical processing and scale-sensitivity using functional MRI (fMRI). As described above, the behavioral work has shown that many listeners cannot discriminate between the major and minor musical modes, even though many people experience music in the major and minor scales as sounding "happy" and "sad," respectively. In fact, 70% of listeners perform near chance in classifying the two modes, while the other 30% are nearly perfect at the task (Chubb et al., 2013). Many musicians and non-musicians alike scored in the high performing group, but many musicians, surprisingly, also fell within the low performing group. Performance in this task correlated with other variants of major and minor tone sequences regardless of presentation rate, suggesting a specialized "scale-sensitivity" in some individuals compared to others that may not be predicted by the number of years of musical experience (Dean and Chubb, 2017; Mednicoff et al., 2018). Thus with such a large proportion of listeners being insensitive, the divide in performance amongst the two populations may be explained through neural differences and through the investigation of scale-sensitivity under fMRI.

Based on the neurobiological literature described above, we predict that cortical differences may first exist within primary auditory belt regions (ie. Heschel's Gyrus and STS), so a localizer scan will be used in the present study to identify these regions of interest (ROIs) for further analysis following previous work (Okada et al., 2010). Having a subject-specific functional localizer will serve to increase the resolution of the analysis, especially considering the anticipation of high inter-subject variability in the locations of functional activations. The hierarchical organization of speech and tones has been evaluated to exist within these primary auditory regions (Okada et al., 2010; Fedorenko et al., 2012; Rogalsky et al., 2011); however, higher musical processing has also been reported to extend to the intraparietal lobule (IPL) (Klein and Zatorre, 2011; Royal et al., 2016). Consistent with this idea, studies that have utilized MVPA for major and minor musical processing have demonstrated significant, common patterns of activity within the right STS and left IPS (Lee et al., 2011; Klein and Zatorre, 2015). Therefore, if cortical differences are not first revealed across the sensitivity levels of listeners with the univariate analysis, then the searchlight procedure may predict significant accuracy peaks in patterns of activity within these two regions of interest for the tone-scramble task that may be differentiable between the two groups of listeners.

## 3.2 Methods

### 3.2.1 Study Participants

Five individuals (N=3, male) participated in the fMRI experiments. Two participants were low-sensitivity listeners, whereas three were high-sensitivity listeners, with sensitivity determined using the tone-scramble task (Chubb et al., 2013). No participant possessed absolute pitch abilities. These participants were selected from a pool of undergraduate and graduate participants at University of California, Irvine, and all of the methods used were approved by the UC Irvine Institutional Review Board.

### 3.2.2 Sound Stimuli

Stimuli in the present study were identical to the stimuli used in the tone-scramble task in Chubb et al. (2013) and were presented to participants over noise-cancelling, high fidelity MR-compatible headphones (ie. OPTIME 1, MR confon, Germany). All stimuli were generated and played using Matlab (Mathworks, Inc.) and the Psychophysics toolbox (Brainard, 1997; Pelli and Vision, 1997; Kleiner et al., 2007).

A given tone-scramble stimulus was composed of a rapid sequence of 32 pure tones: eight copies of both the low and high tonics (G5: 783.99 Hz and G6: 1567.98 Hz), eight of the fifth (D6: 1174.66 Hz), and eight copies of the major third (B5: 987.77 Hz) for major trials or eight copies of the minor thirds (B♭5: 932.33 Hz) for minor trials. Every tone within each tone-scramble had a duration of 65-*milliseconds*, contained 3,250 samples, and was played at a rate of 50,000 samples per second. These tones were presented in a random (or scrambled) order, making each tone-scramble stimulus last 2.08 *seconds*. Every tone was ramped on and off by a raised cosine function with 22 ms rise and fall times to prevent clicking. This musical stimulus allowed us to control for all the elements necessary to create a major or minor key without any rhythmic or timbral effects that are embedded in a typical music selection.

### 3.2.3 fMRI Tasks & Data Acquisition

T1-weighted anatomical scans were first obtained for each participant. The fMRI scanning protocol followed methods used in previous literature (Belin et al., 1999; Klein and Zatorre, 2015). Functional scans consisted of 48 T2*-weighted images covering the entire head (axial slices, 2 x 2 x 2 *mm*, acquired interleaved), acquired using continuous sampling with TR = 1.5*sec*.

The experimental scans proceeded using an event-related design in which listeners were asked to classify the stimulus as major or minor. Participants heard 18 trials in each scan, separated by an inter-stimulus interval 7.5 - 10.5 *sec*, with a mean of 9*sec*. Listeners were first assessed in the task outside of the scanner with trial-by-trial feedback over 100 trials to gain their sensitivity levels and to familiarize themselves to the task; however, no feedback was given to participants in the scanner to avoid the activation of feedback-driven processing.

Primary auditory cortex was localized with two types of stimuli in two different scans. The first used a block design in which 12*sec* of amplitude-modulated broadband noise (at 8 Hz with a modulation depth of 70%) alternated with rest (with scanner noise; see alsoOkada et al. (2010)). The second also used a block design, but instead 12*sec* of major and minor tone-scramble stimuli were alternated with rest in an ABBA design.

### 3.2.4 Imaging analysis

fMRI data were pre-processed using BrainVoyager. All images first underwent slice-timing correction due to interleaved acquisition, followed by motion correction and temporal band-pass filtering with a frequency cut off at 0.01. Due to the collection of only five participants, functional images were co-registered with the subject's native T1-weighted anatomical images for visualization.

### 3.2.5 Localizer Analysis

A general linear model (GLM) with predictors constructed from a boxcar function modeling the duration of the AM noise or tone-scramble stimuli, convolved with a modeled hemodynamic response function Boynton et al. (1996). The amplitude of the BOLD response was estimated as the best fitting beta for the modulated noise. Primary auditory cortex was identified as the region on Heschl's gyrus that was more activated during noise as compared to rest (false discovery rate (FDR; Benjamini and Hochberg (1995)); $q < .001$).

### 3.2.6 MVPA Procedures

Individual trial BOLD amplitude estimates were computed using the least square single (LSS) method for estimating individual trial betas in rapid event-related designs (Mumford et al., 2012). In this approach, individual trial beta amplitudes are iteratively modeled in conjunction with two regressors predicting the hemodynamic response for all the other trial types, split into major and minor scale types. These beta estimates, computed for every voxel in a region-of-interest, were then used for training and testing of a support vector machine classifier trained to label trials as either major or minor. All but one functional run was used for training the classifier ("leave-one-run out"), and then the classifier was tested on the remaining run that was not included in training. Classification accuracy was computed as the average across this five-fold cross-validation.

In an exploratory analysis, support vector machine classification was performed using a searchlight procedure such that the classifier was trained and tested using five-fold cross validation (leave one run out). Voxel patterns for each searchlight were extracted from a sphere with radius of 3 voxels. The average of each accuracy score was assigned to the center voxel of a sphere and stored in an output image for every participant. Significance at the single participant level was measured by whether the classification accuracy differed from chance (50%).

## 3.3 Results

The goal of this project was to use functional neuroimaging to compare the patterns of brain activity in listeners with high vs. low scale-sensitivity. The project aimed to (1) determine where in the brain these regions of activity are, and (2) whether specific regions or distributed patterns of brain activity are differentially activated between listeners high and low in scale-sensitivity. Due to limitations with the COVID-19 pandemic and funding mechanisms, only five participants were collected for this experiment.

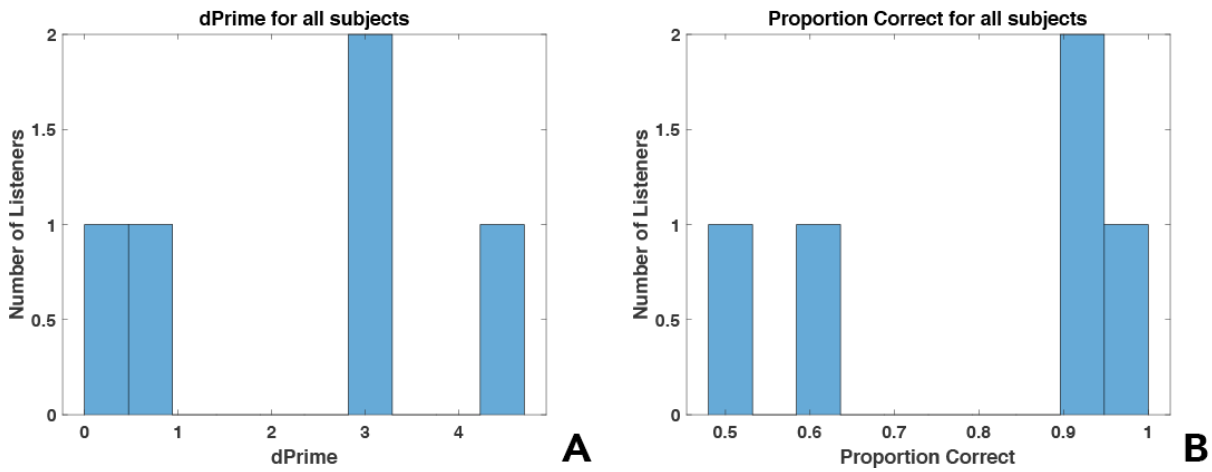


Figure 3.1: *Screened behavioral results in the tone-scramble task (Chubb et al., 2013) for all participants collected for this study.* Figure 3.1a is the distribution of $d'$ results for these participants, and figure 3.1b is the distribution of proportion correct.

### 3.3.1 Behavioral results

All participants were screened to determine their level of sensitivity with the tone-scramble task (Chubb et al., 2013) before participation in the fMRI experiment. Figure 3.1 shows the results for all participants in this task. As shown in Figure 3.1b, two of the listeners achieved a proportion correct between 0.5-0.6, and three of the listeners achieved a proportion correct of 0.9 and above, meaning two listeners had low-sensitivity to this task, and three listeners had high-sensitivity. These same listeners are also shown in Figure 3.1a, where the two listeners with low-sensitivity performed with a $d'$ below 1, and the three listeners with high-sensitivity performed with a $d'$ above 3.

### 3.3.2 GLM analysis

Two localizers, one with amplitude modulated noise and one with major/minor tone-scrambles, were run in addition to the tone-scramble task to determine primary auditory cortices for each participant. A contrast of AM noise > silence revealed no significant clusters, whereas a contrast of tone-scrambles > silence revealed two large clusters bilaterally on the superior

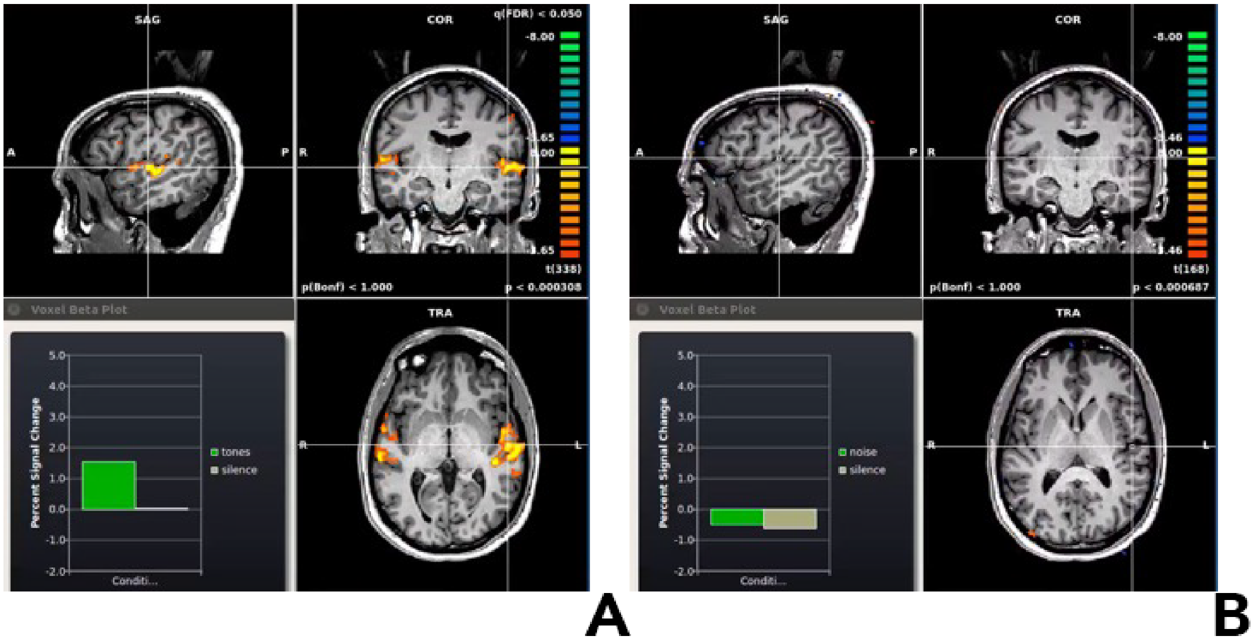


Figure 3.2: *Individual level GLM analyses for one subject which was a high-sensitivity listener.* Figure 3.2a shows the contrast for tone-scramble > silence, and Figure 3.2b shows the contrast for AM noise > silence.

temporal gyrus and Heschl's gyrus, in primary auditory regions. Figure 3.2 shows activation results from one representative participant (from the high-sensitivity group) which was consistently observed for all participants.

There were no statistical differences in neural peaks of activation between the major trials and minor trials anywhere in the brain. Considering only five participants were collected, there were no significant differences in activation between the two groups.

### 3.3.3 MVPA and Searchlight fMRI

Individual level searchlight results revealed variable results across participants. There was no evidence for significant accuracy peaks in auditory or surrounding sensory areas considering only five participants were included in this analysis. When looking at individual level results for each participant, the searchlight analysis revealed regions that contained patterns of

Table 3.1: *Individual level searchlight results for each participant.*

| **Participant Number** | **Type of Listener** | **Region with peak activation** |
|---|---|---|
| Participant 1 | Low–Sensitivity | No significant peaks |
| Participant 2 | High–Sensitivity | Anterior temporal lobe, right lateralized occipital lobe, dorsolateral prefrontal cortex, right lateralized motor cortex |
| Participant 3 | Low–Sensitivity | No significant peaks |
| Participant 4 | High–Sensitivity | Left lateralized motor cortex, medial prefrontal cortex, superior temporal sulcus |
| Participant 5 | High–Sensitivity | Dorsolateral prefrontal precuneus |

activity that classified major vs. minor tone-scrambles. These regions for each participant are shown in Table 3.1.

Figure 3.3 shows the results from the individual level searchlight analysis in native space for one representative participant, which is the same participant shown in Figure 3.2. No significant group level accuracies were computing given the small sample size of the study population.

A MVPA analysis was also run on our five participants. This analysis revealed the classification accuracy for the right and left auditory regions of interest when listening to major vs. minor tone-scrambles. Figure 3.4 contains the results for the MVPA analysis when the cost = 1. While the study sample is too small for statistical analysis, one can observe the trend that the neural patterns in the right hemisphere auditory cortex were able to classify major vs minor patterns more accurately in the three high-sensitivity listeners and not the two low-sensitivity listeners. This finding is suggestive, but not conclusive.

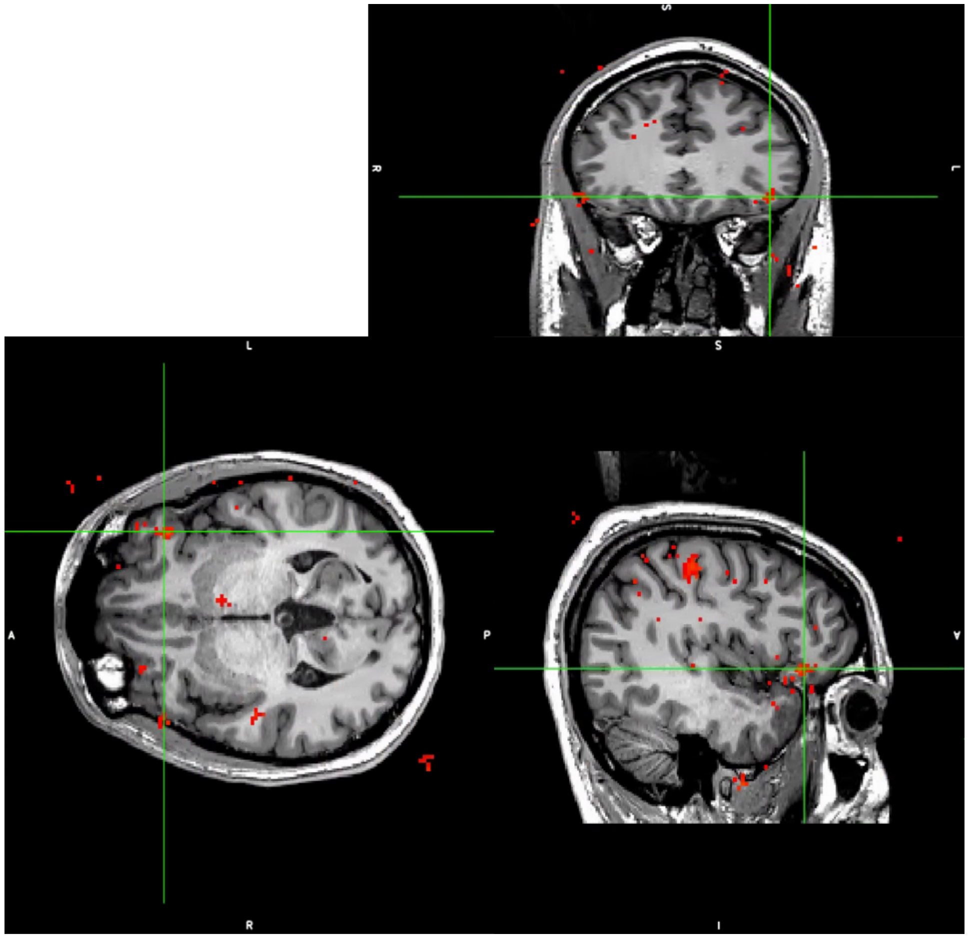

Figure 3.3: *Individual level searchlight results in native space for a high-sensitivity listener.*

## 3.4 Discussion

This experiment sought to uncover neural regions and/or distributed patterns of activity specific to those sensitive and not sensitive to discriminating major and minor musical modes. The behavioral work this study builds upon has shown that 30% of listeners perform near perfect in classifying major and minor modes, whereas the other 70% perform near chance (Chubb et al., 2013). Here, we specifically intended to collect a split sample between the two groups; however, due to limitations of the pandemic, only five participants in total were ultimately collected for the study. Three of these participants were highly sensitive to the task and performed above 90%, and two participants performed near chance.

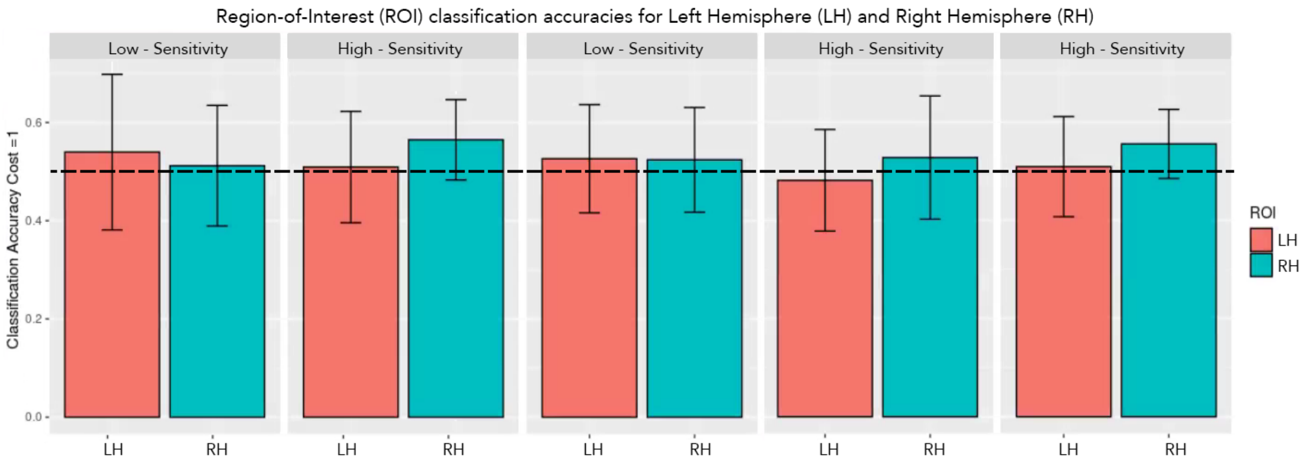


Figure 3.4: *MVPA results when the hyperparameter for each training set was tuned with a cost = 1.* ROI 1 is the left primary auditory region of interest and ROI 2 is the right primary auditory region of interest. The dotted line represents 0.5 chance classification.

The current results do not contain enough power to definitively answer whether listeners high in scale-sensitivity show differences in neural activation from those low in scale-sensitivity. Localizer scans using the tone-scramble stimuli showed two large clusters bilaterally on the superior temporal gyrus and Heschl's gyrus, in primary auditory regions. The searchlight result highlighted individual differences within each participant, but the MVPA results suggested that the pattern of activity in the right hemisphere of the auditory cortex could classify the major vs. minor patterns more accurately for the three high-sensitivity listeners and not the low-sensitivity listeners. Despite this trend, the sample size was too small for statistical analysis.

Although this study collected a small sample of participants, there was one suggestive result from our proposed analyses: those listeners with high-sensitivity to the task have a different pattern of activity in the right primary auditory cortex to classify the major vs. minor tone-scrambles than the listeners with low-sensitivity. Therefore, if this study were to run at a larger scale and collect more participants, then we would be able to determine if listeners high in scale-sensitivity have neural activations that are statistically different from those low in scale-sensitivity. Research that supports this result suggests that the right and left auditory cortices are specialized for spectral and temporal processing, respectively (for review, see Güntürkün et al. (2020)). These results are based on studies that have shown speech sounds to activate neural systems in the left hemisphere, whereas music and tones, or specifically changes in pitch, activate neural systems in the right hemisphere (Zatorre et al., 2002; Tervaniemi and Hugdahl, 2003). On the other hand, additional research suggests that those with specialized musical training may recruit the left hemisphere for additional temporal processing (Ono et al., 2011). Thus, it is unclear to which extent we will see lateralized differences between our two groups of listeners in our study.

In conclusion, there is a different pattern of activity between our listeners that are high and low in scale-sensitivity; however, there was not enough power in our study to statistically determine a difference. Additional studies may seek to address this primary concern.

## 3.5 Acknowledgments.

This work was supported by the UCI Center for Hearing Research project funds. We would like to thank Xiaojue Zhou and Sajjad Torabian for their helpful contributions with the MVPA and searchlight analyses, respectively.

# Bibliography


Adler, S. A., Comishen, K. J., Wong-Kee-You, A. M. B., and Chubb, C. (2020). Sensitivity to major versus minor musical modes is bimodally distributed in young infants. *Journal of the Acoustical Society of America*, 147:3758–3764.

Belin, P., Zatorre, R. J., Hoge, R., Evans, A. C., and Pike, B. (1999). Event-related fmri of the auditory cortex. *Neuroimage*, 10(4):417–429.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1):289–300.

Besson, M., Chobert, J., and Marie, C. (2011a). Language and music in the musician brain. *Language and Linguistics Compass*, 5(9):617–634.

Besson, M., Chobert, J., and Marie, C. (2011b). Transfer of training between music and speech: Common processing, attention, and memory. *Frontiers in Psychology*, 2(94):https://doi.org/10.3389/fpsyg.2011.00094.

Bialystok, E. and DePape, A.-M. (2009). Musical expertise, bilingualism, and executive functioning. *Journal of Experimental Psychology: Human Perception and Performance*, 35(2):565–574.

Blechner, M. J. (1977). Musical skill and the categorical perception of harmonic mode. *Haskins Laboratories Status Report on Speech Perception*, SR-51/52:139–174.

Bonetti, L. and Costa, M. (2019). Musical mode and visual-spatial cross-modal associations in infants and adults. *Musicae Scientiae*, 23(I):50–68.

Boynton, G. M., Engel, S. A., Glover, G. H., and Heeger, D. J. (1996). Linear Systems Analysis of Functional Magnetic Resonance Imaging in Human V1. *The Journal of Neuroscience*, 16(13):4207–4221.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, 10:433–436.

Buss, E., Taylor, C. N., and Leibold, L. J. (2014). Factors affecting sensitivity to frequency change in school-age children and adults. *Journal of Speech Language and Hearing Research*, 57(5):1972–1982.

Chubb, C., Dickson, C. A., Dean, T., Fagan, C., Mann, D. S., Wright, C. E., Guan, M., Silva, A. E., Gregersen, P. K., and Kowalski, E. (2013). Bimodal distribution of performance in discriminating major/minor modes. *Journal of the Acoustical Society of America*, 134(4):3067–3078.

Crowder, R. G. (1984). Perception of the major/minor distinction: I. historical and theoretical foundations. *Psychomusicology*, 4(1/2):3–12.

Crowder, R. G. (1985a). Perception of the major/minor distinction: Ii. experimental investigations. *Psychomusicology*, 5(1/2):3–24.

Crowder, R. G. (1985b). Perception of the major/minor distinction: Iii. hedonic, musical, and affective discriminations. *Bulletin of the Psychonomic Society*, 23(4):314–316.

Cunningham, J. G. and Sterling, R. S. (1988). Developmental change in the understanding of affective meaning in music. *Motivation and Emotion*, 12:399–413.

Curtis, M. E. and Bharucha, J. J. (2010). The minor third communicates sadness in speech, mirroring its use in music. *Emotion*, 10(3):335–348.

Dean, T. and Chubb, C. (2017). Scale-sensitivity: A cognitive resource basic to music perception. *Journal of the Acoustical Society of America*, 142(3):1432–1440.

Fang, R., Ye, S., Huangfu, J., and Calimag, D. P. (2017). Music therapy is a potential intervention for cognition of alzheimer's disease: a mini-review. *Translational neurodegeneration*, 6(1):1–8.

Fedorenko, E., McDermott, J. H., Norman-Haignere, S., and Kanwisher, N. (2012). Sensitivity to musical structure in the human brain. *Journal of Neurophysiology*, 108(12):3289–3300. Publisher: American Physiological Society.

Fujioka, T., Trainor, L. J., Ross, B., Kakigi, R., and Pantev, C. (2004). Musical training enhances automatic encoding of melodic contour and interval structure. *Journal of cognitive neuroscience*, 16(6):1010–1021.

Fujioka, T., Trainor, L. J., Ross, B., Kakigi, R., and Pantev, C. (2005). Automatic encoding of polyphonic melodies in musicians and nonmusicians. *Journal of cognitive neuroscience*, 17(10):1578–1592.

Gagnon, L. and Peretz, I. (2003). Mode and tempo relative contributions to "happy-sad" judgements in equitone melodies. *Cognition and Emotion*, 17(1):25–40.

Gerardi, G. M. and Gerken, L. (1995). The development of affective responses to modality and melodic contour. *Music Perception*, 12(3):279–290.

Güntürkün, O., Ströckens, F., and Ocklenburg, S. (2020). Brain Lateralization: A Comparative Perspective. *Physiological Reviews*, 100(3):1019–1063.

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

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

Heinlein, C. P. (1928). The affective character of the major and minor modes in music. *Comparative Psychology*, VIII(2):101–142.

Hevner, K. (1935). The affective character of the major and minor modes in music. *American Journal of Psychology*, 47:103–118.

Ho, J. and Chubb, C. (2020). How rests and cyclic sequences influence performance in tone-scramble tasks. *Journal of the Acoustical Society of America*, 147:3859–3870.

Janata, P. (2015). Neural basis of music perception. *Handbook of clinical neurology*, 129:187–205.

Johnson, D. M., Watson, C. S., and Jensen, J. K. (1987). Individual differences in auditory capabilities. i. *Journal of the Acoustical Society of America*, 81:427–438.

Juslin, P. N. and Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, 129(5):770–814.

Justice, T. C. and Bharucha, J. J. (2002). Music perception and cognition. In Yantis, S., editor, *Stevens handbook of experimental psychology*, volume 1, pages 453–492. John Wiley, 3rd edition.

Kastner, M. P. and Crowder, R. G. (1990). Perception of the major/minor distinction: Iv. emotional connotations in young children. *Music Perception*, 8(2):189–202.

Kidd, G. R., Watson, C. S., and Gygi, B. (2007). Individual differences in auditory abilities. *Journal of the Acoustical Society of America*, 122:418–435.

Klein, M. E. and Zatorre, R. J. (2011). A role for the right superior temporal sulcus in categorical perception of musical chords. *Neuropsychologia*, 49(5):878–887.

Klein, M. E. and Zatorre, R. J. (2015). Representations of Invariant Musical Categories Are Decodable by Pattern Analysis of Locally Distributed BOLD Responses in Superior Temporal and Intraparietal Sulci. *Cerebral Cortex*, 25(7):1947–1957.

Kleiner, M., Brainard, D., and Pelli, D. (2007). What's new in psychtoolbox-3?

Koelsch, S. and Siebel, W. A. (2005). Towards a neural basis of music perception. *Trends in Cognitive Sciences*, 9(12):578–584.

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

Lee, Y.-S., Janata, P., Frost, C., Hanke, M., and Granger, R. (2011). Investigation of melodic contour processing in the brain using multivariate pattern-based fMRI. *NeuroImage*, 57(1):293–300.

Leek, M. R. and Watson, C. S. (1984). Learning to detect auditory pattern components. *Journal of the Acoustical Society of America*, 76(4):1037–1044.

Leek, M. R. and Watson, C. S. (1988). Auditory perceptual learning of tonal patterns. *Journal of the Acoustical Society of America*, 43(4):389–394.

Macmillan, N. A. and Kaplan, H. L. (1985). Detection theory analysis of group data: estimating sensitivity from average hit and false-alarm rates. detection theory analysis of group data: estimating sensitivity from average hit and false-alarm rates. detection theory analysis of group data: estimating sensitivity from average hit and flase-alarm rates. *Psychological Bulletin*, 98(1):185–199.

Magne, C., Schön, D., and Besson, M. (2006). Musician Children Detect Pitch Violations in Both Music and Language Better than Nonmusician Children: Behavioral and Electrophysiological Approaches. *Journal of Cognitive Neuroscience*, 18(2):199–211.

Mann, D. S. (2014). *Processing Stimuli over Time: Musical Modes and Audiovisual Binding.* PhD thesis, University of California, Irvine.

Marie, C., Delogu, F., Lampis, G., Belardinelli, M. O., and Besson, M. (2011). Influence of musical expertise on segmental and tonal processing in mandarin chinese. *Journal of cognitive neuroscience*, 23(10):2701–2715.

Marie, C., Magne, C., and Besson, M. (2010). Musicians and the metric structure of words. *Journal of cognitive neuroscience*, 23(2):294–305.

Maxwell, J. C. (1855). Experiments on colour, as perceived by the eye with remarks on colour-blindness. *Transactions of the Royal Society of Edinburgh*, XXI, Part II:275–298.

McDermott, J. H., Schemitsch, M., and Simoncelli, E. P. (2013). Summary statistics in auditory perception. *Nature Neuroscience*, 16(4):493–498.

McDermott, J. H. and Simoncelli, E. P. (2011). Sound texture perception via statistics of the auditory periphery: evidence from sound synthesis. *Neuron*, 71:926–940.

Mednicoff, S., Mejia, S., Rashid, J., and Chubb, C. (2018). Many listeners cannot discriminate major vs. minor tone-scrambles regardless of presentation rate. *Journal of the Acoustical Society of America*, 144(4):2242–2255.

Micheyl, C., Delhommeau, K., Perrot, X., and Oxenham, A. J. (2006). Influence of musical and psychoacoustical training on pitch discrimination. *Hearing research*, 219(1):36–47.

Morrill, T. H., Devin, J. D., Dilley, L. C., and Hambrick, D. Z. (2015). Individual differences in the perception of melodic contours and pitch-accent timing in speech: Support for domain-generality of pitch processing. *Journal of Experimental Psychology: General*, 144(4):730–736.

Mumford, J. A., Turner, B. O., Ashby, F. G., and Poldrack, R. A. (2012). Deconvolving bold activation in event-related designs for multivoxel pattern classification analyses. *Neuroimage*, 59(3):2636–2643.

Okada, K., Rong, F., Venezia, J., Matchin, W., Hsieh, I.-H., Saberi, K., Serences, J. T., and Hickok, G. (2010). Hierarchical organization of human auditory cortex: evidence from acoustic invariance in the response to intelligible speech. *Cerebral Cortex*, 20(10):2486–2495.

Ono, K., Nakamura, A., Yoshiyama, K., Kinkori, T., Bundo, M., Kato, T., and Ito, K. (2011). The effect of musical experience on hemispheric lateralization in musical feature processing. *Neuroscience Letters*, 496(2):141–145.

Pallesen, K., Brattico, E., Bailey, C., Korvenoja, A., Koivisto, J., Gjedde, A., and Carlson, S. (2006). Emotion Processing of Major, Minor, and Dissonant Chords. *Annals of the New York Academy of Sciences*, 1060:450–3.

Pantev, C., Oostenveld, R., Engelien, A., Ross, B., Roberts, L. E., and Hoke, M. (1998). Increased auditory cortical representation in musicians. *Nature*, 392(6678):811–814.

Parbery-Clark, A., Skoe, E., Lam, C., and Kraus, N. (2009). Musician enhancement for speech-in-noise. *Ear and hearing*, 30(6):653–661.

Parbery-Clark, A., Strait, D. L., Anderson, S., Hittner, E., and Kraus, N. (2011). Musical experience and the aging auditory system: implications for cognitive abilities and hearing speech in noise. *PLoS One*, 6(5):e18082.

Patel, A. D. (2005). The relationship of music to the melody of speech and to syntactic processing disorders in aphasia. *The Annals of the New York Academy of Sciences*, 1060(1):59–70.

Patel, A. D. (2011). Why would musical training benefit the neural encoding of speech? the opera hypothesis. *Frontiers in psychology*, 2:142.

Patel, A. D. and Iversen, J. R. (2007). The linguistic benefits of musical abilities. *Trends in Cognitive Sciences*, 11(9):369–372.

Patel, A. D., Iversen, J. R., and Rosenberg, J. C. (2006). Comparing the rhythm and melody of speech and music: The case of british english and french. *Journal of the Acoustical Society of America*, 119(5):3034–3047.

Pelli, D. G. and Vision, S. (1997). The videotoolbox software for visual psychophysics: Transforming numbers into movies. *Spatial vision*, 10:437–442.

Peretz, I., Gagnon, L., and Bouchard, B. (1998). Music and emotion: perceptual determinants, immediacy, and isolation after brain damage. *Cognition*, 68:111–141.

Plack, C. J. and Oxenham, A. J. (2005). The psychophysics of pitch. In *Pitch*, pages 7–55. Springer.

Quinn, I. (1999). The combinatorial model of pitch contour. *Music Perception*, 16:439–456.

Rameau, J. P. (1971–orig., 1722). *Treatise on Harmony.* Dover Press, New York.

Rogalsky, C., Rong, F., Saberi, K., and Hickok, G. (2011). Functional Anatomy of Language and Music Perception: Temporal and Structural Factors Investigated Using Functional Magnetic Resonance Imaging. *Journal of Neuroscience*, 31(10):3843–3852. Publisher: Society for Neuroscience Section: Articles.

Romano, A. (2002). *Rising-falling contours in speech: A metaphor of tension-resolution schemes in European musical traditions? Evidence from regional varieties of Italian*, chapter 12, pages 325–337. John Benjamins Publishing.

Royal, I., Vuvan, D. T., Zendel, B. R., Robitaille, N., Schönwiesner, M., and Peretz, I. (2016). Activation in the Right Inferior Parietal Lobule Reflects the Representation of Musical Structure beyond Simple Pitch Discrimination. *PLOS ONE*, 11(5):e0155291. Publisher: Public Library of Science.

Salimpoor, V. N., Benovoy, M., Larcher, K., Dagher, A., and Zatorre, R. J. (2011). Anatomically distinct dopamine release during anticipation and experience of peak emotion to music. *Nature neuroscience*, 14(2):257.

Schmuckler, M. A. (1999). Testing models of melodic contour similarity. *Music Perception*, 16:295–326.

Schmuckler, M. A. (2016). Tonality and contour in melodic processing. In Hallam, S., Cross, I., and Thaut, M., editors, *The Oxford handbook of music psychology*, chapter 15, pages 143–165. Oxford University Press.

Schoenberg, A. (1978–orig. 1922). *Theory of Harmony.* University of California Press.

Schön, D., Magne, C., and Besson, M. (2004). The music of speech: Music training facilitates pitch processing in both music and language. *Psychophysiology*, 41(3):341–349.

Silva, A. E. and Chubb, C. (2014). The 3-dimensional, 4-channel model of human visual sensitivity to grayscale scrambles. *Vision Research*, 101:94 – 107.

Slevc, L. R. (2012). Language and music: sound, structure, and meaning. *WIREs Cognitive Science*, 3(4):483–492. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/wcs.1186.

Strait, D. L., Kraus, N., Parbery-Clark, A., and Ashley, R. (2010). Musical experience shapes top-down auditory mechanisms: evidence from masking and auditory attention performance. *Hearing research*, 261(1-2):22–29.

Strait, D. L., Parbery-Clark, A., Hittner, E., and Kraus, N. (2012). Musical training during early childhood enhances the neural encoding of speech in noise. *Brain and Language*, 123(3):191–201.

Swaminathan, S. and Schellenberg, E. G. (2015). Current emotion research in music psychology. *Emotion review*, 7(2):189–197.

Temperley, D. and Tan, D. (2013). Emotional connotations of diatonic modes. *Music Perception*, 30(3):237–257.

Tervaniemi, M. and Hugdahl, K. (2003). Lateralization of auditory-cortex functions. *Brain Research Reviews*, 43(3):231–246.

Tymoczko, D. (2011). *A Geometry of Music–Harmony and Counterpoint in the Extended Common Practice*. Oxford University Press.

Victor, J. D., Conte, M. M., and Chubb, C. (2017). Textures as probes of visual processing. *Annual review of vision science*, 3:275–296.

Viemeister, N. F. (1979). Temporal modulation transfer functions based upon modulation thresholds. *Journal of the Acoustical Society of America*, 66(5):1364–1380.

Virtala, P. and Tervaniemi, M. (2017). Neurocognition of major-minor and consonance-dissonance. *Music Perception: An Interdisciplinary Journal*, 34(4):387–404.

Waz, S. and Chubb, C. (2019). Evidence of a single neural mechanism underlying scale-sensitivity. *Society for Music Perception and Cognition (SMPC) Conference, New York*, Poster:https://osf.io/x8v5n.

Zatorre, R. J., Belin, P., and Penhune, V. B. (2002). Structure and function of auditory cortex: music and speech. *Trends in Cognitive Sciences*, 6(1):37–46.

Zuk, J., Benjamin, C., Kenyon, A., and Gaab, N. (2014). Behavioral and neural correlates of executive functioning in musicians and non-musicians. *PLoS ONE*, 9(6):e99868. doi:10.1371/journal.pone.0099868.