UNIVERSITY OF CALIFORNIA,
IRVINE

Searching for the key to musical scale-sensitivity through rhythm, speech, and pitch

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Cognitive Sciences

by

Joselyn Ho

Dissertation Committee:
Professor Charlie Chubb, Chair
Professor Gregory Hickok
Professor Virginia Richards

2021

Dedication

Shout-out to my favorite musicians: Lights, BTS, and Twenty One Pilots.

Contents

List of Figures

List of Tables

Acknowledgements

This dissertation and countless other projects would not have been successful without the work of the Tonaliteam research assistants: Jesus Aguilar, Joey Almaraz, Olivia Capizzi, Yasmin Delavar, Amanda Emsais, Cristy Gonzales, Karen Gonzalez, Melissa Huynh, Suyeon Hwang, Nellie Kwang, Nickole Livas, Elvira Lopez, Christopher Ngov, Nguyen Pham, Juergen Riedelsheimer, Anastasiya Toroptseva, Alfonso Valenzuela, and Luis Zambrano. Every contribution, no matter how small, has made an impact on my research journey and I am excited for what the future holds for you. Thank you so much for your time, initiative, and the community that you bring to the lab!

To my advisor, Charlie Chubb: Thank you for recognizing my potential when I applied to join the lab five years ago and investing your time into my graduate journey. You have been welcoming and patient with me since day 1, and I appreciate your guidance and support as I explored different projects to build confidence in my abilities.

To my committee, Greg Hickok and Ginny Richards. Greg: Thank you for welcoming me into your lab, providing alternate perspectives on my projects, and connecting me with valuable opportunities such as the autism study and the online tone-scramble game with Harvard Music Lab. Ginny: Thank you for teaching me so much about EEG and hearing, training me to read papers more efficiently and critically, and getting me to engage more with background literature through journal club!

Thank you to Haleh Farahbod for helping me so much with research logistics and IRB paperwork, and for making the Auditory & Language Neuroscience Lab feel like a family!

Thank you to Kourosh Saberi and Ramesh Srinivasan for taking the time to mentor me through various avenues of my research journey. Thank you to Amy Bauer for serving on my advancement committee and giving valuable feedback. Thank you to Barbara Sarnecka for the skills that I learned through your writing workshop and for helping me recognize that rejections are worth celebrating.

Thank you to my funding sources. A huge thanks to the UC Irvine Center for Hearing Research for promoting the hearing sciences in all disciplines and providing me the training to conduct quality research. (The projects in this dissertation were funded by NIH Training Grant No. T32-DC010775.)

To my Tonaliteam, Solena Mednicoff and Sebastian Waz: Somehow the pandemic brought us closer together and I am so grateful for it! Solena: Thank you for your wonderful friendship and mentorship over these years. I really appreciate that no matter how busy you get, you enthusiastically help me out with whatever I'm struggling with. Sebastian: I learned so much about data science from you! Thank you for sharing your expertise in stats, coding, and obscure music recommendations.

To my best friends, Ai Ohno and Kelly Park: You have been with me for all of my major life events and I am eternally grateful for your love and support in everything I do!

To my parents: Thank you for supporting me through my career even when it is new and unfamiliar, and welcoming me home anytime to eat delicious food and do my laundry.

To Josephine and Meow Pie: Hi Jie Jie and Meow Pie!!!!!

Vita

Joselyn Ho

EDUCATION

Doctor of Philosophy in Cognitive Sciences **2021**
University of California, Irvine *Irvine, CA*

Masters of Science in Cognitive Neuroscience **2019**
University of California, Irvine *Irvine, CA*

Bachelor of Science in Cognitive Science **2016**
University of California, Los Angeles *Los Angeles, CA*

TEACHING EXPERIENCE

Instructor **Summer 2020**
University of California, Irvine *Irvine, CA*

Teaching Assistant **2016–2021**
University of California, Irvine *Irvine, CA*

Teaching Assistant **Summer 2015–2016**
Johns Hopkins Center of Talented Youth *Saratoga Springs, NY & Los Angeles, CA*

WORK EXPERIENCE

Data Science Intern **July-Sept 2018**
Northrop Grumman *Albuquerque, NM*

FELLOWSHIPS AND GRANTS

Associate Dean Fellowship **2021**
University of California, Irvine *Irvine, CA*

Center of Hearing Research Training Grant **2018-2020**
University of California, Irvine *Irvine, CA*

Research Travel Grant **2019**
University of California, Irvine *Irvine, CA*

Rapid Grant Award **2016**
University of California Music Experience Research Community Initiative *CA*

Psychology Departmental Honors **2015-2016**
University of California, Los Angeles *Los Angeles, CA*

PUBLICATIONS

- **Ho, J.**, Chubb, C. How rests and cyclic sequences influence performance in tone-scramble tasks. The Journal of the Acoustical Society of America, 147(6), 3859-3870.

- Bufford, C. A., Thai, K. P., **Ho, J.**, Xiong, C., Hines, C. A., & Kellman, P. J. (2016). Perceptual Learning of Abstract Musical Patterns: Recognizing Composer Style. Proceedings of the 14th International Conference on Music Perception and Cognition, 8-12.

Forthcoming manuscripts

- **Ho, J.**, Hickok, G., & Chubb, C. (in prep). Musical Mode Sensitivity Predicts Sensitivity to Speech Independently of Musical Training.

- **Ho, J.**, Mann, D., Hickok, G., & Chubb, C. (in prep). Many listeners have roved pitch-comparison thresholds above a quarter-tone; very few can discriminate major from minor tone-scrambles.

CONFERENCE PRESENTATIONS

Talks

- **Ho, J.** & Chubb, C. (2022, May). Musical Scale-Sensitivity Predicts Sensitivity to Speech Prosody Independently of Musical Training. Talk to be presented at the Expression, Language, & Music Conference, Hartford, CT.

- **Ho, J.** (2019, Sept). The influence of rhythmic and sequential structure on musical scale-sensitivity. Talk presented at the So Cal Hearing Conference, Irvine, CA.

Posters

- **Ho, J.** & Chubb, C. (2019, Aug). The influence of rhythmic and sequential structure on classifying major vs. minor tone-scrambles. Poster presented at the Biennial Meeting of the Society for Music Perception and Cognition, New York City, NY.

- **Ho, J.**, Hickok, G., & Chubb, C. (2018, Nov). Musical sensitivity correlates with pitch production ability in speech. Poster presented at the 59th Annual Meeting of the Psychonomic Society, New Orleans, LA.

- **Ho, J**. & Chubb, C. (2017, July). Modeling the effect of scale impurities on tonality perception. Poster presented at the Society for Music Perception and Cognition Conference, San Diego, CA.

- Bufford, C. A., Thai, K. P., **Ho, J.**, Xiong, C., Hines, C. A., & Kellman, P. J. (2016, July). Perceptual Learning of Abstract Musical Patterns: Recognizing Composer Style. Poster presented at the 14th International Conference on Music Perception and Cognition, San Francisco, CA.

- Hines, C. A., **Ho, J.**, Xiong, C., Bufford, C., Thai, K. P., & Kellman, P. J. (2016, May). Perceptual Learning Intervention Improved Abstract Musical Pattern Recognition. Poster presented at the UCLA Psychology Undergraduate Research Conference, Los Angeles, CA.

- **Ho, J.**, Xiong, C., Bufford, C., Thai, K. P., Chun, J., & Kellman, P. J. (2015, May). Perceptual Learning of Musical Abstract Patterns. Poster presented at the UCLA Psychology Undergraduate Research Conference, Los Angeles, CA, and the Stanford Undergraduate Psychology Conference, Stanford, CA.

- **Ho, J.**, Shiboski, E. M., Xiong, C. Y., Bufford, C., Thai, K. P., & Kellman, P. J. (2014, May). Perceptual Learning of Abstract Patterns in Music. Poster presented at the Stanford Undergraduate Psychology Conference, Stanford, CA, and the Berkeley Interdisciplinary Research Conference, Berkeley, CA.

Abstract of the Dissertation

Searching for the key to musical scale-sensitivity through rhythm, speech, and pitch

By

Joselyn Ho

Doctor of Philosophy in Cognitive Sciences

University of California, Irvine, 2021

Professor Charlie Chubb, Chair

This dissertation investigates the sources of musical scale-sensitivity, or the sensitivity to musical mode. In Chapter 1, I introduce the concept of scale-sensitivity, the tone-scramble "3-task" paradigm that can be used to measure this skill, and the open questions surrounding the bimodal distribution of scale-sensitivity in the general population. In Chapter 2, I investigate whether the temporal structure of tone-scramble stimuli influences scale-sensitivity. By manipulating the speed and the grouping of tones in the stimuli, I find that inserting regular, brief rests into the tone sequences heightens sensitivity to musical mode, and that specific note sequences can strongly bias listeners to perceive a stimulus as one type over another. In Chapter 3, I investigate whether scale-sensitivity is related to sensitivity to speech prosody. I find evidence that scale-sensitivity and speech sensitivity may depend on shared processing resources that are largely unaffected by musical training. In Chapter 4, I explore the relationship between scale-sensitivity and pitch-difference threshold by testing listeners in variations of a pitch comparison task. I find that having a pitch-difference threshold below 50 cents on a roved pitch comparison task is required to achieve high scale-sensitivity. Finally, in Chapter 5, I discuss the implications of these findings on music and emotion perception and present next steps in continuing to understand the source of scale-sensitivity.

Chapter 1

Introduction

Music is found in every human culture examined so far and connects with us deeply. One of several ways that music can convey emotions is through mode, referring to particular combinations of musical notes. For example, on average, across listeners, music in the major (Ionian) mode tends to sound "happy" while music in the minor (Aeolian) mode tends to sound "sad" (Cunningham & Sterling, 1988; Gagnon & Peretz, 2003; Gerardi & Gerken, 1995; Heinlein, 1928; Hevner, 1935; Kastner & Crowder, 1990; Leaver & Halpern, 2004; Peretz, Gagnon, & Bouchard, 1998; Temperley & Tan, 2013). As a result of this striking qualitative difference, the major and minor scales have come to play a central role in western music. This difference in perceived emotional quality might imply that listeners can naturally tell apart major from minor melodies; however, many listeners, including musicians, struggle to differentiate the two types (Halpern, 1984; Halpern, Bartlett, & Dowling, 1998; Leaver & Halpern, 2004).

Sensitivity to major vs. minor musical modes appears to be bimodally distributed across listeners. This effect was first made apparent in the results of Crowder (1985b). In a task replicating Blechner (1977), listeners strove to classify triadic chords as major vs. minor. Each stimulus was a 300-ms triad from the equal-tempered scale in either root position (note-order from low to high: tonic, third, fifth) or first inversion (note-order from low to high:

third, fifth, tonic). Across trials, the tonic varied randomly (between 6 notes). The fifth of the triad was 7 semitones above the tonic, and the third of the triad was 1 to 9 logarithmic steps between the minor and major third, relative to the tonic. The task was to classify the triad according to whether the third was closer to the minor vs. the major third. Despite the small sample size (19 subjects), Crowder (1985b) observed that the psychometric functions (relating the third in the triad to the probability that the subject responded "major") fell into two distinct groups. The psychometric function was either very steep, suggesting that the listener was highly sensitive to the major-minor difference, or flat, suggesting that the listener had little or no sensitivity to the major-minor difference. Only three listeners fell in the middle between these two extremes.

A bimodal distribution of performance has also been observed in a major-minor "tone-scramble" task (e.g., Chubb et al. (2013)). In the following chapters of this dissertation, the tone-scramble "3-task" paradigm (described below) is implemented to investigate listeners' musical sensitivity.

1.1 Tone-scramble "3-Task"

In the basic version of the tone-scramble task (the "3-task"), the participant listens to rapid (923 BPM), randomly-ordered sequences of pure tones and attempts to classify each as major (happy) or minor (sad). Feedback is given after every trial. The stimuli, called "tone-scrambles," contain 8 each of the notes G_5, D_6 and G_6 (to establish G as the tonic); in addition, major tone-scrambles contain 8 B_5's (degree 3 of the G major scale) whereas minor tone-scrambles contain 8 $B\flat_5$'s (degree 3 of the G minor scale). Data pooled from multiple studies (Chubb et al., 2013; Dean & Chubb, 2017; Mednicoff, Mejia, Rashid, & Chubb, 2018; Ho & Chubb, 2020) shows that 70% of listeners perform near chance on this task, while the remaining 30% perform almost perfectly (Fig. 1.1).

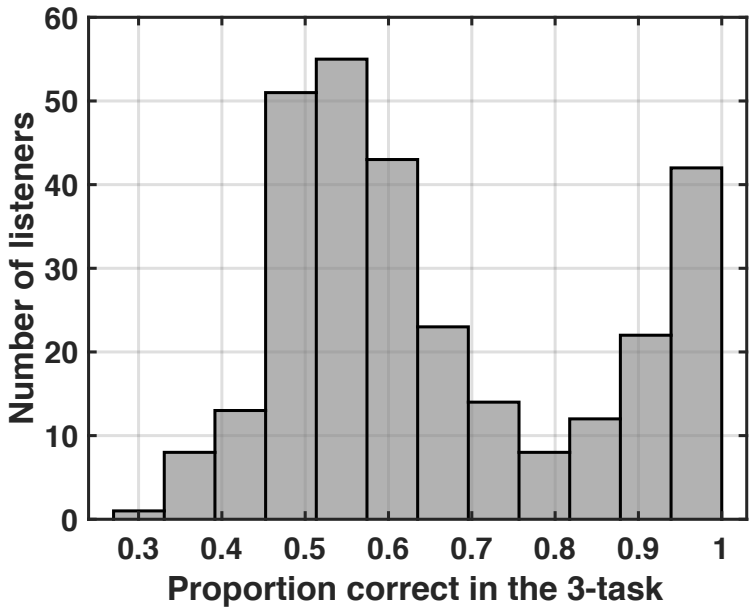

Figure 1.1: Histogram of performance in the 3-task combining results from Chubb et al. (2013), Dean and Chubb (2017), and Mednicoff et al. (2018).

1.1.1 Musical training

The striking bimodal distribution of performance on the 3-task suggests that high-performers possess capabilities that low-performers lack. Interestingly, skill on this task is only weakly related to one's musical background (measured in years of formal musical training). Fig. 1.2 plots 3-task-d' against years-musical-training for the same 293 listeners whose results are plotted in Fig. 1.1. The correlation of 0.35 is highly significant showing that listeners with musical training tend to perform better than those without musical training. However, this correlation is due mainly to a large group of listeners with no musical training who perform poorly. Notably, there are also many low-performers with many years of musical training (aqua ellipse) as well as other high-performers who have few years of musical training, suggesting that years-musical-training is neither necessary nor sufficient for high performance on the 3-task. Perhaps the positive correlation between years-musical-training and 3-task-d' arises because listeners with high scale-sensitivity are more likely to seek out musical training than listeners with low scale-sensitivity. This idea is supported by the finding that 6-month-old infants show the same distribution of performance in the 3-task as adults

(Adler, Comishen, Wong-Kee-You, & Chubb, 2020), suggesting that a listener's level of scale-sensitivity may be hereditary or formed early in life.

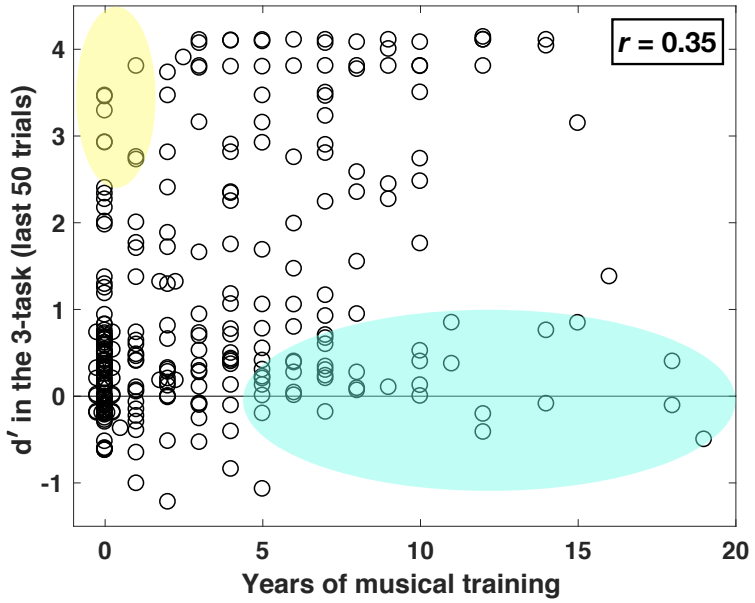

Figure 1.2: Scatterplot of d' achieved in the 3-task as a function of years of musical training, combining results from Chubb et al. (2013), Dean and Chubb (2017), and Mednicoff et al. (2018). Aqua ellipse (lower right) indicates listeners with many years of musical training who perform poorly in the 3-task. Yellow ellipse (upper left) indicates listeners with little or no musical training who perform well in the 3-task.

Indeed, evidence from both behavioral and genetic studies demonstrate that other factors besides musical training (e.g., pre-existing abilities) may contribute to individual differences in musical abilities (Correia et al., 2020; Kragness, Swaminathan, Cirelli, & Schellenberg, 2020; Hambrick & Tucker-Drob, 2015; Mosing, Madison, Pedersen, Kuja-Halkola, & Ullén, 2014). For example, a study by Kragness et al. (2020) found a positive relationship between musical training and musical ability in young children, but not when prior musical ability is held constant. Additionally, musical ability assessed at an earlier time point predicted the duration of subsequent musical training over the next 5 years, providing further evidence that the connection between musical training and musical ability is not a straightforward causal relationship. Genetic studies investigating twins have also found genetic influences on musical accomplishment, and these effects were enhanced in individuals receiving musical

training (Hambrick & Tucker-Drob, 2015). Therefore, musical training alone cannot account for individual differences on the tone-scramble task – high-performers may possess a processing resource that confers heightened sensitivity to the difference between the major-minor stimuli.

1.2 Current Work

The studies described in the following chapters explore the source of major-minor sensitivity on the tone-scramble task and possible influences. Chapter 2 addresses whether this sensitivity is affected by repetition and temporal grouping. If sensitivity depends on temporal structure, then introducing manipulations such as rests and cyclic sequences into the stimuli may heighten the sensitivity of listeners to the differences between the stimulus types. Chapter 3 investigates the relationship between major-minor sensitivity and speech prosody perception. Substantial evidence supports a close link between music and speech perception, particularly in the processing of pitch cues; therefore, there may be overlap between the computational resources required for processing speech prosody and musical mode. Chapter 4 investigates whether pitch-difference threshold is a contributing factor to musical mode sensitivity. Chapter 5 discusses the implications of these findings and suggests potential studies that can be conducted to further our understanding of the tone-scramble task phenomenon.

Chapter 2

Temporal structure of tone-scrambles influences sensitivity to mode

2.1 Abstract

When classifying major versus minor tone-scrambles (random sequences of pure tones), most listeners (70%) perform at chance while the remaining listeners perform nearly perfectly. The current study investigated whether inserting rests and cyclic sequences into the stimuli could heighten sensitivity in such tasks. In separate blocks, listeners classified tone-scramble variants as major versus minor ("3" task) or fourth versus tritone ("4" task). In three "Fast" variants, tones were played at 65 ms/tone as a continuous, random stream ("FR"), or with a rest after every fourth tone ("FRwR"), or as a repeating sequence of four tones with a rest after every fourth tone ("FCwR"). In the "Slow" variant, tones were played at 325 ms/tone in random order. In both the "3" and "4" tasks, performance was ordered from best to worst as follows: FRwR > FR > FCwR > Slow. Post-hoc analysis revealed that performance was suppressed in the Slow and FCwR task-variants due to a powerful bias inclining listeners to respond "major" or "fourth" ("minor" or "tritone") if the 4-note sequence defining the stimulus ended on a high (low) note. Overall, the results indicate that inserting regular rests

into random tone sequences heightens sensitivity to musical mode.

2.2 Introduction

A bimodal distribution in performance is observed in a task requiring listeners to classify major vs. minor "tone-scrambles" (Chubb et al., 2013; Dean & Chubb, 2017; Mednicoff et al., 2018). In the basic major-minor task (the "3-task"), each tone-scramble contains 32, 65-ms tones including 8 copies each of the notes G_5, D_6, G_6 (to establish G as the tonic of each stimulus), and a target note. The target note in major tone-scrambles is B_5 (the third degree of the G major scale), and the target note in minor tone-scrambles is $B\flat_5$ (the third degree of the G minor scale). On each trial, the listener hears a single tone-scramble and attempts, with feedback, to classify it as major or minor.

Dean and Chubb (2017) tested listeners in a range of tasks akin to the 3-task but using different pairs of target notes. For example, in the "4-task", the target notes that differentiated the two stimulus types were C_6 (fourth scale degree of both the G major and G minor scales) and $D\flat_6$ (a tritone above G_5, included in neither the G major nor G minor scales). The results were well-described by a bilinear model which proposes that sensitivity of listener k to the difference between the two types of tone-scramble stimuli used in task t ($d'_{k,t}$) is determined by the listener's amount of scale-sensitivity (R_k) and the strength with which scale-sensitivity facilitates performance in task t (F_t). Specifically,

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

Dean and Chubb (2017) concluded that performance in all of the tasks t used in their study was determined predominantly by a single processing resource, R. Like in Chubb et al. (2013), listeners' performance took the form of a bimodal distribution: 70% percent of listeners possessed levels of R near zero which yielded near-chance performance in all tasks, while 30% of listeners possessed levels of R that yielded higher performance. R also

facilitated different tasks with different strengths. Since the target notes used in most of the tasks were unrelated to the difference between the major vs. minor scales, Dean and Chubb (2017) concluded that R confers general sensitivity to variations in scale with a fixed tonic. They therefore called R "scale-sensitivity," proposing that listeners possess different levels of this resource which determines their ability to discriminate tone-scramble types.

It has also been shown that performance in the 3-task does not improve if stimuli are presented more slowly (Mednicoff et al., 2018). Listeners in this study were tested in major-vs-minor tone-scramble classification tasks in which the stimuli were played at different rates. In the slowest condition (which we refer to below as the Slow-3 task), each tone-scramble contained 4, 520-ms tones: G_5, D_6, G_6, and a target note ($B\flat_5$ or B_5). Listeners who performed poorly in any condition performed poorly across all conditions. Surprisingly, performance was the worst in the slowest condition, and listeners' responses were strongly affected by the order of the four tones in each tone-scramble, regardless of whether the stimulus was major or minor. Specifically, even though note-order is irrelevant to the task, listeners were biased to respond "major" if a tone-scramble ended on a high note. In the slowest condition, the responses of more than half of all listeners were influenced more strongly by this shared bias than they were by whether the target note was $B\flat_5$ vs. B_5.

The current study explores the following question: Does scale-sensitivity depend on temporal structure? If so, then perhaps scale-defined properties (e.g., majorness vs. minorness) can be made more legible by introducing rhythmic and/or sequential structure into tone-scramble stimuli. It has long been recognized that chunking isolated pieces of information together can improve processing (Miller, 1956). Music typically comprises phrases, or subunits of a longer melody, that can be defined through temporal structure. For example, sequences of tones that occur within a rhythmic pattern are better remembered than sequences that span rhythmic patterns (Dowling, 1973). Further, sequences with regularly-occurring rests are better remembered than sequences that occur as a continuous stream, and the tones between each rest tend to be recalled or forgotten as a unit (Deutsch, 1982).

Therefore, introducing temporal structure into tone-scramble stimuli may heighten the sensitivity of listeners to the differences between the stimulus types. In the current study, we focused on the effects of rests and cyclic sequences (i.e. repeating sets of 4 tones).

We also sought to more deeply explore the relationship between scale-sensitivity and the note-order-specific response biases that tend to subvert performance in the Slow-3 task of Mednicoff et al. (2018). There was a strong, shared tendency to classify the Slow-3 stimuli as "major" if they ended on a high note (especially the high tonic). We speculated that this effect was provoked by the suggestion (made in the response prompt presented visually after each trial) to classify stimuli as major if they sounded "happy" and minor if they sounded "sad." The prevalence of the ending-on-a-high-note bias (across all listeners other than those with very high scale-sensitivity) suggests that, perhaps for some reason rooted in language-processing, the stimuli in the Slow-3 task of Mednicoff et al. (2018) naturally sound happier if they end on a high note vs. a low note.

To test this possibility, we included four task conditions that might provoke sequence-specific biases. In two of these tasks, the stimuli differ in majorness vs. minorness (as in the study of Mednicoff et al. (2018)). In the other two tasks, the stimuli differ in a quality that might be described as harmoniousness vs. dissonance. The target notes in the major-minor tasks are the third scale degrees of the major and minor scales. The target notes in the other task are the fourth scale degree (which is in both the major and minor scales) and the tritone (which is in neither).

2.3 Method

All methods were approved by the UCI Institutional Review Board.

2.3.1 Participants

Ninety-eight listeners participated in this study and were all undergraduate students at the University of California, Irvine, with self-reported normal hearing. Sixty-nine listeners reported having at least one year of formal musical training. The mean number of years of musical training across all 98 listeners was 4.5 (standard deviation: 4.9). All listeners received course credit for participating in the study.

2.3.2 Stimuli

The experiment used eight stimulus variants with two types each (determined by which target note it contained), for a total of 16 stimulus types (Table 2.1). Stimuli were tone-scrambles, which are sequences of pure tones comprising equal numbers of a target note **T** plus three other notes from the standard equal-tempered chromatic scale: G_5 (783.99 Hz), D_6 (1174.66 Hz), and G_6 (1567.98 Hz). In the "3-task" variants, the target note **T** was $B\flat_5$ (932.33 Hz) for "low-target" (Minor) stimuli or B_5 (987.77 Hz) for "high-target" (Major) stimuli. In the "4-task" variants, **T** was C_6 (1046.50 Hz) for "low-target" (Fourth) or $D\flat_6$ (1108.73 Hz) for "high-target" (Tritone) stimuli. Thus, in all tasks, the high target note was a semitone higher in pitch than the low target note.

Stimuli in the six "Fast" task variants contained twenty, 65-ms tones. Tones in the FR-3 and FR-4 ("Fast Random") tasks were presented in a continuous stream. Tones in the FRwR-3 and FRwR-4 ("Fast Random with Rests") tasks were presented as five bursts of four tones. Each burst contained a random sequence of the notes G_5, D_6, G_6 and **T**, and bursts were separated by 130-ms rests. Tones in the FCwR-3 and FCwR-4 ("Fast Cyclic with Rests") tasks were presented in five repeating bursts of the same sequence of four tones (one each of G_5, D_6, G_6 and **T**), and bursts were separated by 130-ms rests.

The stimuli in the Slow-3 and Slow-4 tasks comprised one each of the notes G_5, D_6, G_6 and **T**, played in random order at 325 ms per tone.

In all tasks, each individual tone per stimulus was windowed by a raised cosine function

with a 22.5-ms rise time.

2.3.3 Procedure

At the start of the experiment, listeners completed a brief survey to report (among other information) their number of years of musical training.

Listeners were then tested in each of the FR-3, FRwR-3, FCwR-3, Slow-3, FR-4, FRwR-4, FCwR-4, and Slow-4 tasks. Task order was randomly generated for each listener.

At the start of each task, the listener heard eight example stimuli labeled as either "Type 1" or "Type 2." In 3-task variants, Type 1 corresponded to high-target (major) stimuli; Type 2 corresponded to low-target (minor) stimuli. In 4-task variants, Type 1 corresponded to low-target (fourth) stimuli; Type 2 corresponded to high-target (tritone) stimuli. These distinctions were not explicitly told to the listener. Then, on each trial, the listener heard a single stimulus and strove to judge which type was presented by entering "1" or "2" for their response. Correctness feedback was printed to the screen after each trial, and proportion correct was given at the end of each block.

Each task consisted of two blocks of 48 trials. Stimulus type (high- vs. low-target) was determined randomly on each trial. To shorten the experiment duration, the number of trials

Table 2.1: *The temporal and sequential properties of the stimuli used in the 8 tasks.* The number in the task name (i.e., 3 or 4) indicates that task's set of target notes ($B\flat_5/B_5$ and $C_6/D\flat_6$, respectively).

Tasks	Number of tones	Tone duration	Rests	Order
FR-3, FR-4	20	65 ms	no	random
FRwR-3, FRwR-4	20	65 ms	yes	random
FCwR-3, FCwR-4	20	65 ms	yes	cyclic
Slow-3, Slow-4	4	325 ms	no	random

in the first block of the FR-3, FR-4, FRwR-3, and FRwR-4 tasks was reduced to 24 for the last 70 listeners. For the basic analysis (Sec. 2.4), we computed listeners' d' values from the second block of trials and treated the first block of trials as practice, as has been done in previous tone-scramble studies (Chubb et al., 2013; Dean & Chubb, 2017; Mednicoff et al., 2018). In analyzing the sequence-specific biases that occur in the FCwR-3, FCwR-4, Slow-3, and Slow-4 tasks (Sec. 2.6), we used both blocks of 48 trials to increase statistical power.

The experiment took place 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/s, and listeners wore JBL Elite 300 noise-cancelling headphones with volume adjusted to their comfort level.

2.4 Results

Listeners' d' values, our basic dependent measure, were computed using the last 48 trials of each task. The first block of 24 or 48 trials per task was treated as practice. If a listener was tested on n high-target (low-target) stimuli over the course of the last 48 trials and responded correctly on all of them, then the probability of a correct response was adjusted to $\frac{n-0.5}{n}$ (as suggested by Macmillan and Kaplan (1985)). This implies that d' values around 4.1 correspond to near-perfect performance on all 48 trials of a task.

Comparisons of the d' values achieved by listeners in different tasks suggest that tasks differed in difficulty. Table 2.2 lists the results of paired samples t-tests of the null hypothesis that the mean value of d' is equal for two tasks. Some of the main trends revealed by this table are: (1) performance in each of the FRwR-3, FR-3, FCwR-3, Slow-3 tasks is significantly better than performance in the corresponding 4-task; and (2) for each of $n = 3, 4$, performance in the Slow-n task is significantly worse than in the FRwR-n, FR-n, and FCwR-n tasks.

Performance showed a significant tendency to improve across tasks. Specifically, we

computed the linear trend L_k in the vector of eight d' values achieved by each listener k across the eight tasks in the order in which the listener was tested. The mean value of the L_k's was 0.18. A 1-tailed t-test of the null hypothesis that the true mean was 0 yielded $t_{97} = 2.30$, $p = 0.012$.

2.4.1 Bilinear model results

Using the bilinear model (Eq. 2.1), we estimated F_t and R_k values. Following the analysis procedure of Mednicoff et al. (2018), we set the constraint that

$$\sum_{\text{tasks } t} F_t \;=\; 8 \qquad \text{(where 8 is the number of tasks).} \tag{2.2}$$

This constraint has convenient properties. First, if all tasks are equally facilitated by R, then F_t will be 1 for all tasks. Second, Eq. 2.2 makes R_k the average value of d' achieved by

Table 2.2: *Results from paired samples t-tests of d' achieved by all listeners in each pair of tasks.* The values shown are t-statistics for 2-tailed tests of the null hypothesis that the mean value of d' is equal for the two tasks. All t-statistics have 97 degrees of freedom. * $p < 0.05/28 = 0.0018$ (Bonferroni correction).

Task	FR-3	FRwR-3	FCwR-3	Slow-3	FR-4	FRwR-4	FCwR-4	Slow-4
FR-3	—	1.14	-0.34	5.52*	4.59*	1.40	4.20*	5.24*
FRwR-3		—	0.74	5.71*	5.76*	2.70	5.53*	6.27*
FCwR-3			—	6.41*	5.76*	2.08	6.04*	6.41*
Slow-3				—	0.23	3.35*	0.15	1.79
FR-4					—	4.56*	0.05	1.72
FRwR-4						—	4.14*	4.47*
FCwR-4							—	1.72
Slow-4								—

listener k across all 8 task-variants.

The estimated values of F_t for all tasks t are displayed in Fig. 2.1. As suggested by the d' results, and consistent with the results of Dean and Chubb (2017), F_t is higher for each 3-task variant t (t = FRwR-3, FR-3, FCwR-3, and Slow-3) than it is for the corresponding 4-task variant.

For $n = 3, 4$, $F_{FRwR-n} > F_{FR-n} \approx F_{FCwR-n} > F_{Slow-n}$. In particular, F_{Slow-n} is much lower than F_t for each of the "Fast" conditions (t = FRwR-n, FR-n, FCwR-n). This is consistent with the finding of Mednicoff et al. (2018) that listeners perform worse when tone-scrambles are played more slowly.

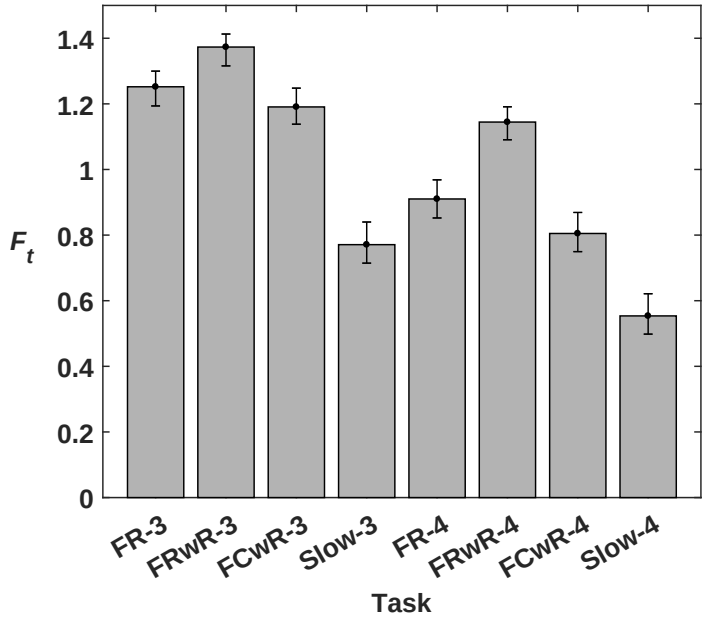

Figure 2.1: *Estimated values of F_t for the eight tasks.* Error bars are 95% Bayesian credible intervals. Figure reproduced from [Ho, J., & Chubb, C. (2020). How rests and cyclic sequences influence performance in tone-scramble tasks. The Journal of the Acoustical Society of America, 147(6), 3859-3870.], with the permission of AIP Publishing.

The left panel of Fig. 2.2 displays the histogram of R_k estimated for the 98 listeners k. Similar to the histogram of R values observed by both Dean and Chubb (2017) and Mednicoff et al. (2018), this histogram is positively-skewed with most listeners possessing R values near 0. This histogram does not appear bimodal. However, as seen in the right panel

of Fig. 2.2, the histogram of proportion correct that these listeners would be predicted to achieve in the FR-3 task (assuming they used optimal criteria) yields the bimodal distribution of performance that is typically observed for this task-variant (Chubb et al., 2013; Dean & Chubb, 2017; Mednicoff et al., 2018).

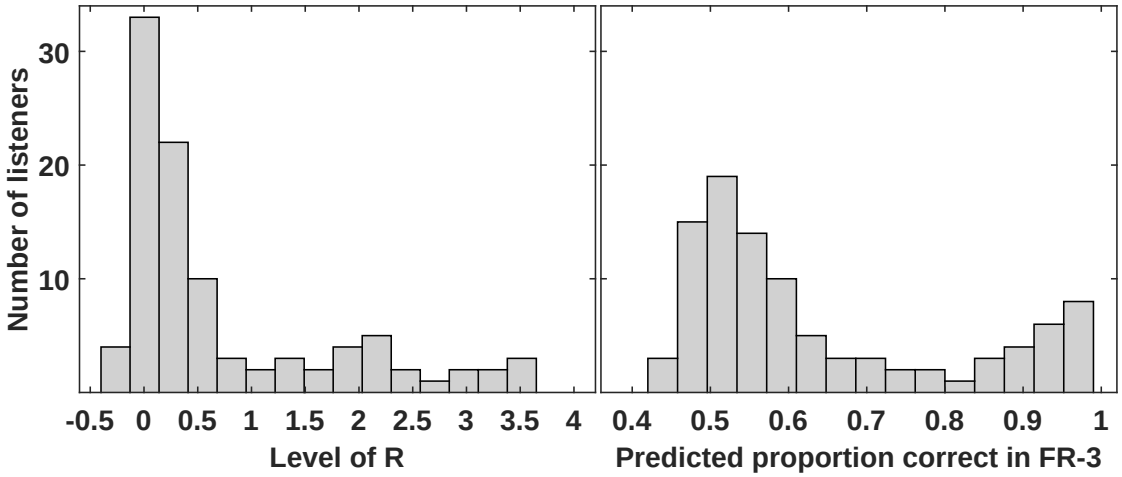

Figure 2.2: Left panel: Histogram of estimated R levels for the 98 listeners. This histogram is positively-skewed with most listeners possessing R values near 0. Right panel: Histogram of predicted proportions correct in the FR-3 task corresponding to the R levels in the left panel. This histogram appears bimodal although the histogram of R levels does not. Figure reproduced from [Ho, J., & Chubb, C. (2020). How rests and cyclic sequences influence performance in tone-scramble tasks. The Journal of the Acoustical Society of America, 147(6), 3859-3870.], with the permission of AIP Publishing.

The results are well-described by the bilinear model. Fig. 2.3 plots the estimates of $d'_{k,t}$ for each listener k in each task t against the values predicted by the bilinear model, and a strong relationship is observed. The bilinear model accounts for 73.8% of the variance in the values of $d'_{k,t}$ for the 98 listeners across the eight tasks.

2.4.2 Relationship with music training

Fig. 2.4 plots each listener's R against his-or-her self-reported years of musical training, showing a significant correlation of 0.364 (p<0.01). In the group of 37 listeners with at least five years of musical training, 21 listeners had R values below 1. Three of the listeners in

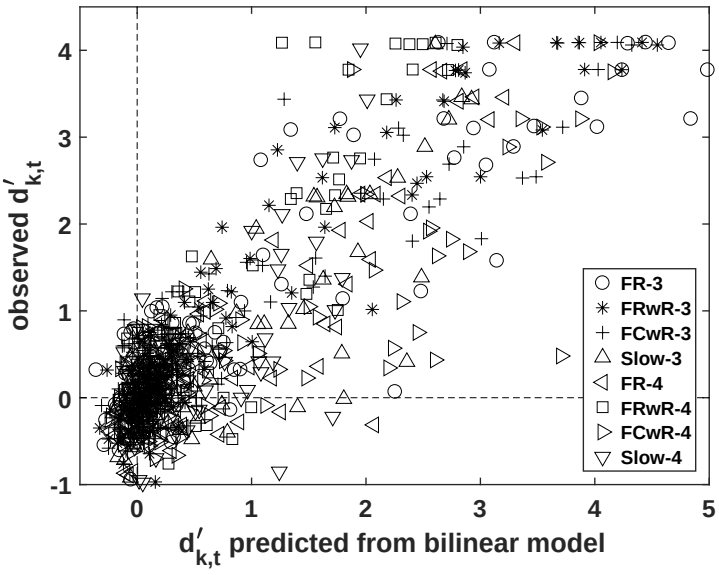

Figure 2.3: *Scatterplot of observed d′ values against estimated d′ values from the bilinear model.* Each point corresponds to a given listener k in a given task t. Tasks are indicated by the symbols displayed in the legend. Figure reproduced from [Ho, J., & Chubb, C. (2020). How rests and cyclic sequences influence performance in tone-scramble tasks. The Journal of the Acoustical Society of America, 147(6), 3859-3870.], with the permission of AIP Publishing.

this group of 21 had at least 15 years of musical training.

The highest R value attained by the 35 listeners with fewer than two years of musical training was 2.2. Among the six listeners who attained R values above 3, four listeners had at least five years of musical training. Therefore, listeners with high values of R tend to have more years of musical training, which follows the pattern observed by Dean and Chubb (2017) and Mednicoff et al. (2018).

2.5 Discussion

The current study explored the degree to which scale-sensitivity (Dean & Chubb, 2017) is modulated by basic variations in temporal structure, which were implemented through periodic rests and cyclic note sequences.

Similar to the results of Dean and Chubb (2017) and Mednicoff et al. (2018), performance

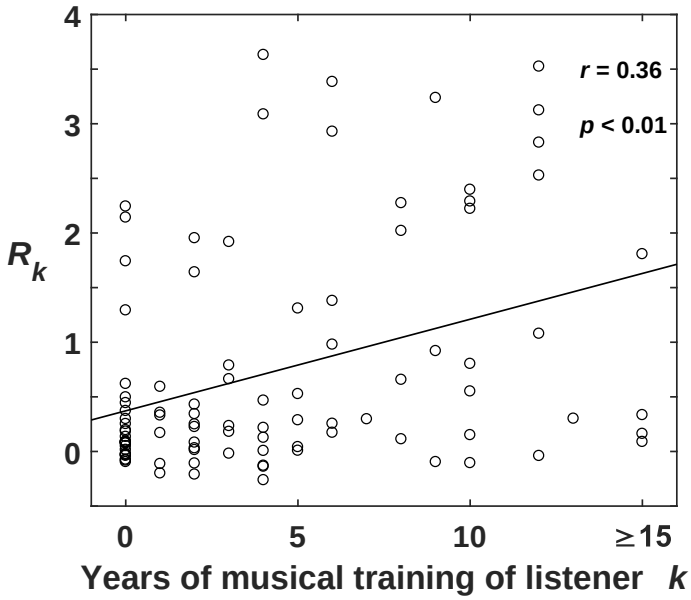

Figure 2.4: *The relationship between musical training and R_k.* Figure reproduced from [Ho, J., & Chubb, C. (2020). How rests and cyclic sequences influence performance in tone-scramble tasks. The Journal of the Acoustical Society of America, 147(6), 3859-3870.], with the permission of AIP Publishing.

.

was well-described by the bilinear model (Eq. 2.1) across all listeners k in all tasks t, implying that performance on the tasks in this study is primarily determined by a single processing resource. The current study is linked to Mednicoff et al. (2018) and Dean and Chubb (2017) in the shared use of the FR-3 task. This commonality suggests that a single processing resource (called "scale-sensitivity" by Dean and Chubb (2017)) underlies performance in all three studies.

We also note that for each of $n = 3, 4$, $F_{FRwR-n} > F_{FR-n}$. The current study does not clearly determine the basis of this effect because the stimuli in the FRwR-n task differ from those in the FR-n task in two ways. First, the FRwR-n task stimuli contained a rest between each burst of four notes. Second, each burst contained one each of the notes G_5, D_6, G_6, and **T**. Either or both of these features may have contributed to the heightened performance in the FRwR-n task, compared to in the FR-n task.

For each of $n = 3, 4$, $F_{FRwR-n} > F_{FCwR-n}$. In the FCwR-n task, the five bursts in each

stimulus repeat the same randomly-ordered sequence of the four notes G_5, D_6, G_6, and **T**. By contrast, in the FRwR-n task, each of the five bursts in each stimulus contains a random sequence of the four notes. In Sec. 2.6, we present evidence suggesting that the heightened difficulty of the FCwR-n task vs. the FRwR-n task stems from systematic response biases associated with individual sequences of the four notes G_5, D_6, G_6, **T**. These biases operate more strongly to subvert performance in the FCwR-n task than they do in the FRwR-n task.

2.6 Note-Order Effects

In this section, we focus exclusively on the FCwR-3, FCwR-4, Slow-3, and Slow-4 tasks because a given stimulus in any of these tasks is completely determined by a single sequence of four notes. (This is not true in any of the other tasks.) In this section, we investigate whether specific permutations of these notes influence listeners' responses. Following Mednicoff et al. (2018), we first use a "Descriptive" model to capture the detailed structure in the data. We then show that the results from the Descriptive model can be captured by a much simpler "Note-function-biased" model.

2.6.1 Notation

We refer to the notes G_5, $B\flat_5$, B_5, C_6, $D\flat_6$, D_6, and G_6 by their respective pitch height values (1, 4, 5, 6, 7, 8, 13), which represent the notes' locations in the chromatic scale starting at G_5.

In each of the FCwR-3, Slow-3, FCwR-4 and Slow-4 tasks, a given stimulus corresponds to a particular 4-note sequence. We refer to individual notes as "pips;" the symbol S refers to the four-pip sequence that determines a stimulus in a task. For $t = 1, 2, 3, 4$, $S(t)$ is the note assigned to pip t. Also in each task, all stimuli are constructed from a set $Notes = \{1, T^-, T^+, 8, 13\}$, where T^- denotes the lower of the two target notes and T^+

denotes the other (a semitone higher in pitch). In the FCwR-3 and Slow-3 Tasks, $T^- = 4$ and $T^+ = 5$, and in the FCwR-4 and Slow-4 Tasks, $T^- = 6$ and $T^+ = 7$. The symbol **T** refers to the target note in a stimulus. We call a stimulus with target note T^- (T^+) a low-target (high-target) stimulus.

A permutation of the four symbols "1", "8", "13", and "**T**" produces a "note-order." Substituting "T^-" for **T** in a given note order Q yields a symbol string corresponding to a stimulus S_Q^-. Substituting T^+ for **T** yields a symbol string corresponding to a stimulus S_Q^+.

2.6.2 Modeling framework

In the general modeling framework, a listener k computes an internal statistic for each stimulus S and compares it to a criterion η_k which is fixed across all trials in a given task.

Let $M_{k,S}$ be the expectation of this internal statistic. Then both of the models we consider assume

$$\text{Response of listener } k \text{ to stimulus } S = \begin{cases} \text{"high-target"} & \text{if } M_{k,S} + X > \eta_k \\ \text{"low-target"} & \text{if } M_{k,S} + X < \eta_k \end{cases}, \qquad (2.3)$$

where X is a standard normal random variable.

2.6.3 Fitting procedures

In the following sections, we fit the Descriptive and Note-function-biased models to the data in the FCwR-3, Slow-3, FCwR-4, and Slow-4 tasks. We derive maximum likelihood estimates of all parameters. In addition, we use a Bayesian fitting procedure to derive credible intervals around these estimates. Specifically, we assume a jointly uniform prior distribution with wide bounds on all model parameters. Then, using Markov chain Monte Carlo simulation, we extract a sample of vectors from the posterior joint density characterizing the parameters. In the figures in this section, line markers show maximum likelihood estimates of parameters,

and error bars give the 0.025 and 0.975 quantiles of posterior, marginal parameter densities.

2.6.4 The Descriptive model

For listeners k with high values of R_k, we expect $M_{k,s}$ to depend strongly on the value of

$$\tau_S = \begin{cases} 1 & \text{if } S \text{ is a high-target stimulus,} \\ -1 & \text{if } S \text{ is a low-target stimulus.} \end{cases} \tag{2.4}$$

As Mednicoff et al. (2018) discovered in the Slow-3 task, many listeners also exhibit shared, systematic, S-dependent response biases. In the Descriptive model, these biases are captured by free parameters β_S corresponding to all 48 possible stimuli S (24 note-orders $\times$ 2 target notes). The Note-function-biased model (described below) uses a simplified rule to predict the β_S values.

Both the Descriptive and Note-function-biased models assume that

$$M_{k,S} = f_\tau(R_k)\tau_S + f_\beta(R_k)\beta_S \tag{2.5}$$

where the function $f_\tau(R)$ reflects the strength with which τ_S influences the response of a listener with scale-sensitivity R, and the function $f_\beta(R)$ reflects the strength with which β_S influences the response of a listener with scale-sensitivity R.

In order to uniquely specify the descriptive model, we impose several constraints on the parameters; these are described in Sec. 2.6.6.

2.6.5 The Note-function-biased model

The Note-function-biased model describes a simple theory of how the β_S's are computed. For a task with low and high target notes T^- and T^+, let $Notes = \{1, T^-, T^+, 8, 13\}$. Under the Note-function-biased model, there exist functions $f_{\text{note}} : Notes \to \mathbb{R}$ and $f_{\text{pip}} : \{1, 2, 3, 4\} \to$

$\mathbb{R}$ such that

$$\beta_S = \sum_{t=1}^{4} f_{\text{note}}(S(t)) f_{\text{pip}}(t), \tag{2.6}$$

where $S(t)$ is the note occurring at pip t in the sequence defining S.

The "Pitch-height-biased" model used by Mednicoff et al. (2018) is a special case of the Note-function-biased model in which

$$f_{\text{note}} \;=\; f_{PH}(n) \;=\; n - M_{Notes} \qquad \text{for all } n \in Notes, \tag{2.7}$$

where M_{Notes} is the mean of the notes $n \in Notes$.

In order to uniquely specify the Note-function-biased model, we impose several constraints on the parameters; these are described in Sec. 2.6.6.

2.6.6 Model Constraints

We impose several constraints on the parameters for each of the two models.

The stimulus-specific biases β_S from the Descriptive model are constrained as follows:

$$\sum_S \beta_S = 0 \qquad \text{and} \qquad \frac{1}{48}\sum_S \beta_S^2 = 1, \tag{2.8}$$

where each sum is over all 48 stimuli S. The first constraint prevents β_S values from trading off with the threshold values η_k. The second constraint prevents β_S values from trading off with f_β, and also enables comparison of their magnitudes to those of the τ_S values (which also satisfy $\frac{1}{48}\sum_S \tau_S^2 = 1$).

Following Mednicoff et al. (2018), we forced the functions $f_\tau(R)$ and $f_\beta(R)$ from the Descriptive model 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 study.

The parameters of the Descriptive model are the 48 β_S values, the six values each of f_τ and

f_β, and the 98 η_k values. Therefore, taking into account the two degrees of freedom sacrificed by imposing the constraints of Eq. 2.8 on the β_S values, the model absorbs $48+12+98-2 = 156$ degrees of freedom.

The reader will note that the Note-function-biased model is under-constrained. For example, for any choice of the functions f_{note} and f_{pip} in the model, and any non-zero scalar α, if we replace f_{note} and f_{pip} with $\hat{f}_{\text{note}} = \alpha f_{\text{note}}$ and $\hat{f}_{\text{pip}} = \frac{f_{\text{pip}}}{\alpha}$, the new model will yield exactly the same predictions. To uniquely determine model parameters, we must specify the relative signs and amplitudes of f_β, f_{note}, and f_{pip}. We impose particular constraints to facilitate comparison of results from the FCwR-3, Slow3, FCwR-4, and Slow-4 tasks and also from Mednicoff et al. (2018).

To make the results from the Note-function-biased model comparable to those of the Pitch-height-biased model of Mednicoff et al. (2018), we constrain f_{note} to sum to 0 and also to have the same sum of squares as f_{PH} (Eq. 2.7). To ensure that the β_S values that result from Eq. 2.6 will satisfy Eq. 2.8, f_{pip} is constrained to sum to 0, and scaled to make β_S's satisfy the right side of Eq. 2.8. In addition, $f_{\text{pip}}(4)$ is constrained to be positive (which makes the f_{pip}'s from all four tasks similar in form). Finally, the sum of f_β taken across sextiles $3, 4, 5$ of the R distribution is constrained to be positive (which makes the f_β's from all four tasks similar in form).

Thus, the total number of degrees of freedom absorbed by the Note-function-biased model is 115: f_{pip} uses 2 degrees of freedom; f_{note} uses 3; the η_k's use 98; and each of f_τ and f_β uses 6.

2.6.7 Modeling results

Instead of considering directly the parameters β_S for all sequences S, it is useful to focus instead on the equivalent, alternative parameters

$$\mu_Q = \frac{\beta_{S_Q^+} + \beta_{S_Q^-}}{2} \quad \text{and} \quad \delta_Q = \frac{\beta_{S_Q^+} - \beta_{S_Q^-}}{2} \tag{2.9}$$

for all note-orders Q. For a given note-order Q, μ_Q reflects the bias injected by the note-order Q regardless of target note, and δ_Q reflects the difference in influence exerted by T^+ vs T^- in the context of Q. (Note that $\beta_{S_Q^+} = \mu_Q + \delta_Q$, and $\beta_{S_Q^-} = \mu_Q - \delta_Q$.)

Across the four tasks, many of the μ_Q values estimated from the descriptive model (plotted as black circles in Fig. 2.5) deviate significantly from 0, confirming that note-order exerts strong influence on stimulus-specific biases regardless of the target note. By contrast, very few of the δ_Q values estimated from the descriptive model (black circles in Fig. 2.6) deviate significantly from 0 suggesting that note order does not strongly influence the relative influence exerted by T^+ vs T^-.

In Figs. 2.5 and 2.6, the gray triangles plot the results from the Note-function-biased model. As described in the Appendix, f_{note} uses only 3 degrees of freedom and f_{pip} uses only 2 as a result of the model constraints. Thus the Note-biased-function model uses only five degrees of freedom to account for all of the 24 μ_Q's and 24 δ_Q's. As reflected by the descriptive model fit, it captures the overall structure in the data remarkably well. In particular, for each of the four tasks, a likelihood ratio test (Hoel, Port, & Stone, 1971; Wilks, 1944) fails to reject the nested Note-function-biased model in favor of the fuller Descriptive model. Under the null hypothesis, the test statistic X is chi-square distributed with 41 degrees of freedom. For the FCwR-3 task, $X = 44.7$, $p = 0.32$; for the Slow-3 task, $X = 26.7$, $p = 0.96$; for the FCwR-4 task, $X = 47.4$, $p = 0.23$; and for the Slow-4 task, $X = 54.8$, $p = 0.07$.

Fig. 2.7 plots the functions f_τ (black) and f_β (gray). As expected, f_τ increases with R_k. For all four tasks, the Descriptive model estimates of f_β are, on average, significantly above 0 and roughly equal for listeners with scale-sensitivity levels in sextiles 4, 5, and 6. Thus, the biases reflected by the black circles in Fig. 2.5 operate with roughly equal strength across all listeners with scale-sensitivity greater than ≈ 0.3.

The relative influence of the biases β_S on the judgments of our listeners varies strongly across the four tasks. Consider a listener k with R_k in the sixth sextile. In the FCwR-3 task, $f_\tau(R_k)$ is around 6 times greater than $f_\beta(R_k)$; this implies that the identity of the

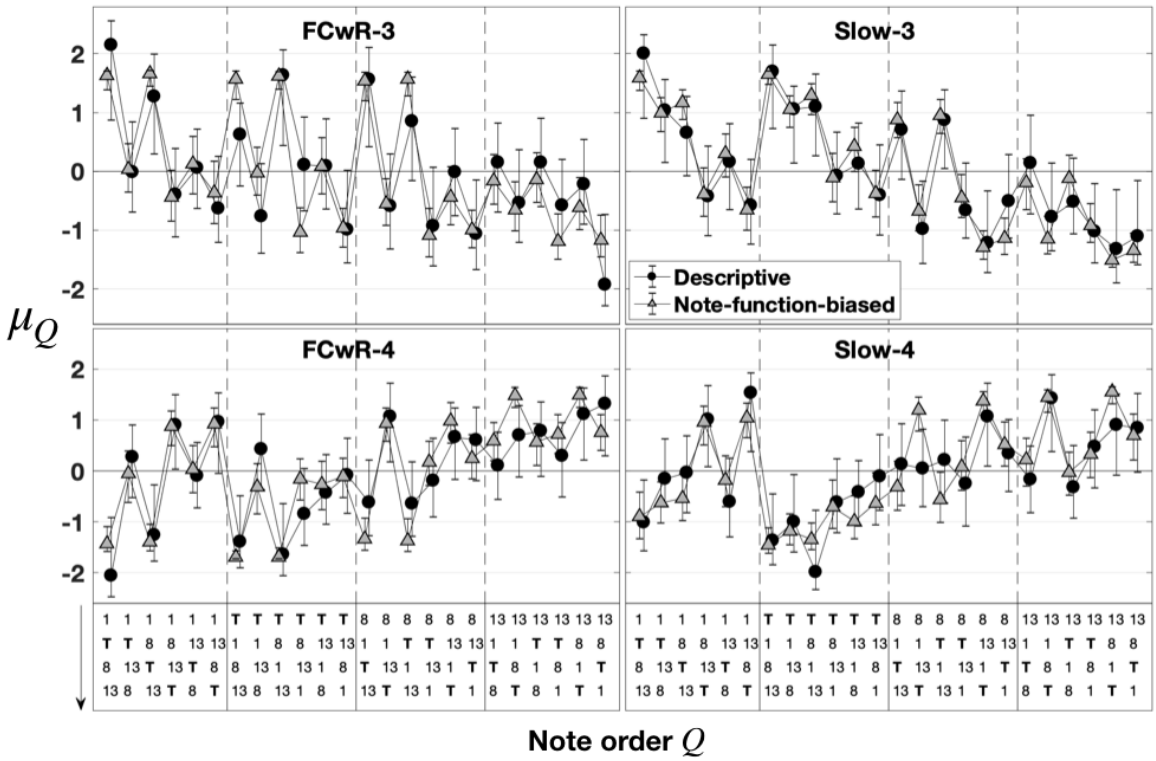

Figure 2.5: *The note-order-specific biases μ_Q for the 24 note-orders Q in each of the four tasks.* The note-order Q of a given stimulus S is represented along the horizontal axis, running downward. Values estimated from the Descriptive model (Note-function-biased model) are plotted in black circles (gray triangles). Markers show maximum likelihood estimates; error bars show 95% Bayesian credible intervals. Figure reproduced from [Ho, J., & Chubb, C. (2020). How rests and cyclic sequences influence performance in tone-scramble tasks. The Journal of the Acoustical Society of America, 147(6), 3859-3870.], with the permission of AIP Publishing.

target note exerts roughly 6 times more influence on the response of this listener than do the sequence-specific biases. At the other extreme, however, in the Slow-4 task, $f_\tau(R_k)$ is only around twice as great as $f_\beta(R_k)$. In the Slow-3 task, $f_\tau(R_k)$ is around 4 times greater than $f_\beta(R_k)$, and in the FCwR-4 task, $f_\tau(R_k)$ is around 3.5 times greater than $f_\beta(R_k)$.

The left panel of Fig. 2.8 plots the temporal weighting function f_{pip} for all four tasks. As described in the appendix, f_{pip} is constrained to sum to 0 and to have $f_{\text{pip}}(4) > 0$; thus, all four functions rise up similarly.

The differences between the 3- and 4-task variants are concentrated in the note weights function f_{note} (right panel of Fig. 2.8). In each of the FCwR-3 and Slow-3 tasks, f_{note}

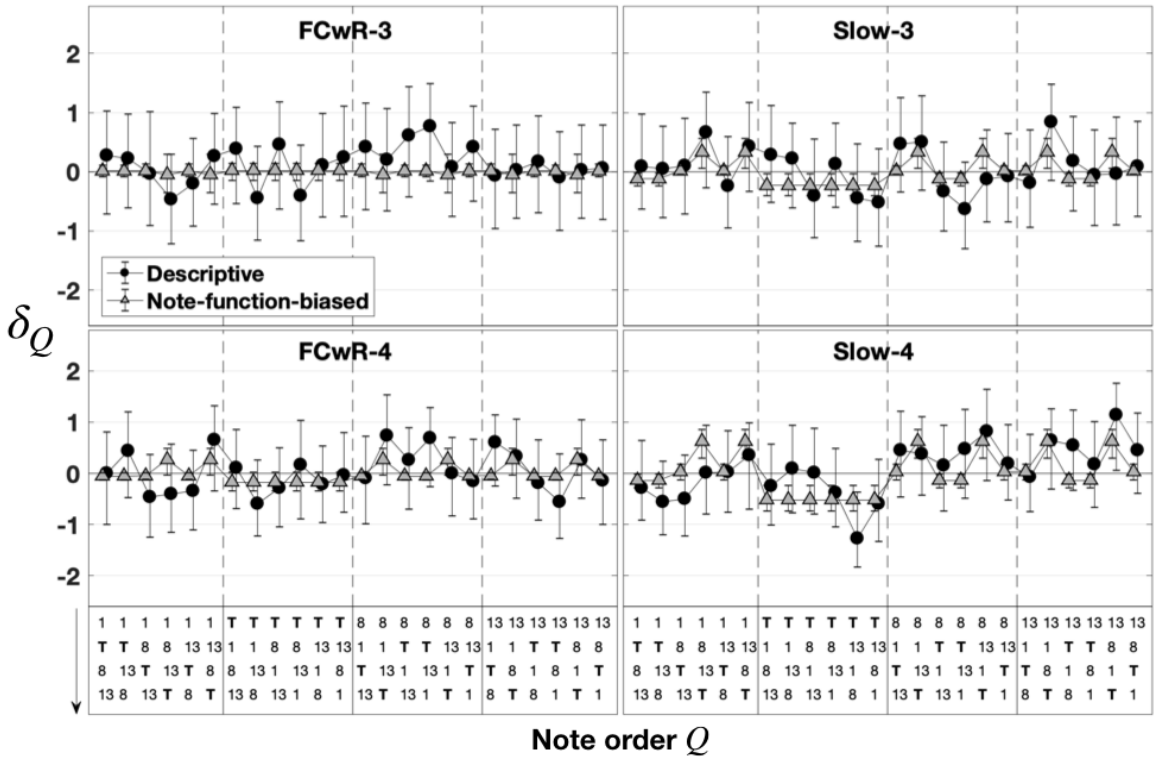

Figure 2.6: *The note-order-specific differences in influence exerted by T^+ vs T^- for the 24 note-orders Q in each of the four tasks.* The note-order Q of a given stimulus S is represented along the horizontal axis, running downward. Values estimated from the Descriptive model (Note-function-biased model) are plotted in black circles (gray triangles). Markers show maximum likelihood estimates; error bars show 95% Bayesian credible intervals. Figure reproduced from [Ho, J., & Chubb, C. (2020). How rests and cyclic sequences influence performance in tone-scramble tasks. The Journal of the Acoustical Society of America, 147(6), 3859-3870.], with the permission of AIP Publishing.

is similar to f_{PH} (Eq. 2.7). This is not true for the FCwR-4 and Slow-4 tasks: f_{note} reaches its maximum at T^+ and descends to its minimum at note 13. It should also be noted that $f_{\text{note}}(T^-) \approx f_{\text{note}}(T^+)$ in the FCwR-3 task. In the other three tasks, however, $f_{\text{note}}(T^-) < f_{\text{note}}(T^+)$.

2.6.8 Discussion of note-order effects

The note-order effects first observed by Mednicoff et al. (2018) were unanticipated and mysterious. The judgment required in the Slow-3 task depends only on which of the two target notes (pitch height values 4 or 5) occurs in the stimulus; the order of the notes is

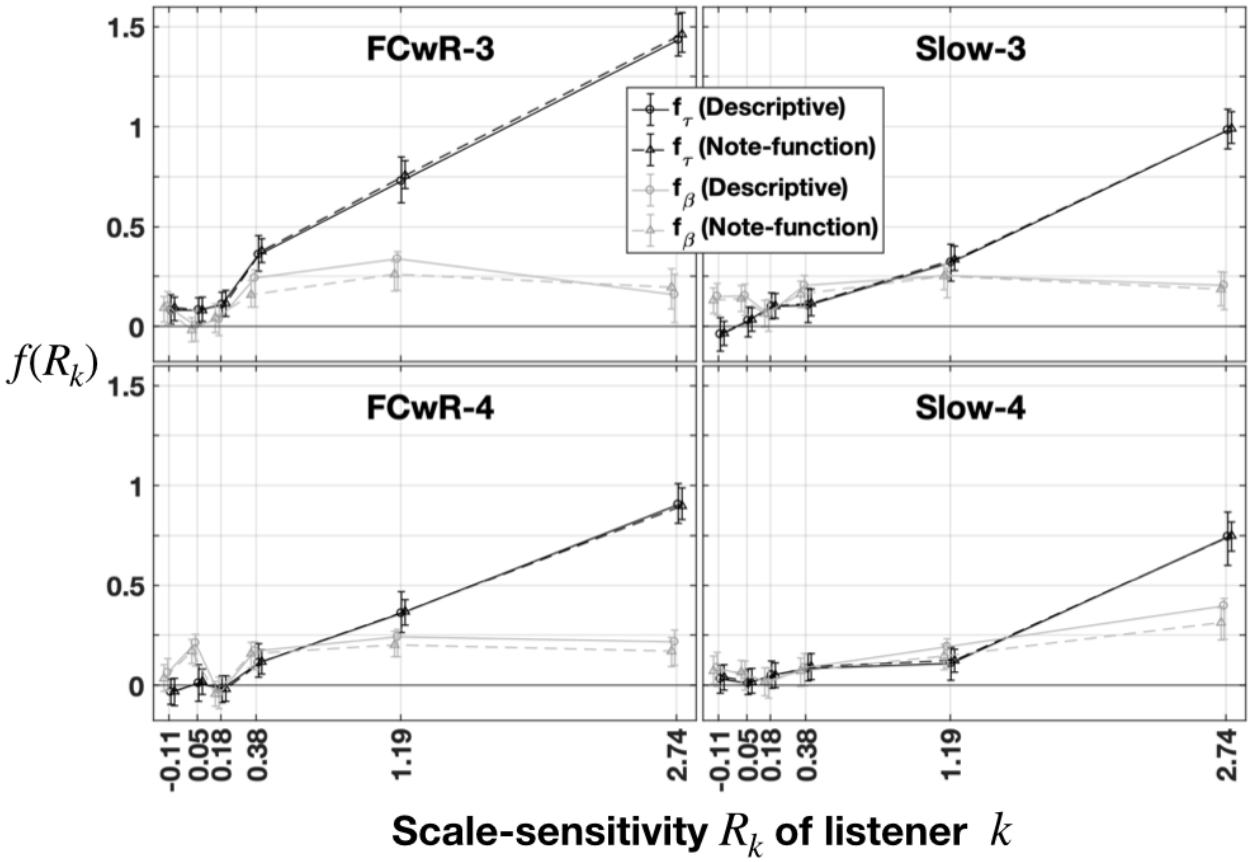

Figure 2.7: *The functions f_τ and f_β in the four tasks.* The values of R_k along the horizontal axis are the mean values of the six sextiles of R_k observed across the 98 listeners. f_τ is plotted in black; f_β is plotted in gray. Solid (dashed) lines show the fits from the Descriptive (Note-function-biased) model. Error bars are 95% Bayesian credible intervals. Figure reproduced from [Ho, J., & Chubb, C. (2020). How rests and cyclic sequences influence performance in tone-scramble tasks. The Journal of the Acoustical Society of America, 147(6), 3859-3870.], with the permission of AIP Publishing.

irrelevant. Nonetheless, the listeners' judgments were strongly influenced by shared biases that depend on the note-order of the stimulus. Mednicoff et al. (2018) accounted for their results in terms of the Pitch-height-biased model, which proposed that listeners' responses to a given stimulus S are influenced by a bias β_S according to Eq. 2.6, with f_{note} equal to the function f_{PH} (Eq. 2.7) and f_{pip} similar in form to the functions plotted in Fig. 2.8.

The current experiment sought to broaden our understanding of these biases by probing two questions:

1. *How do the biases depend on the target notes used in a given task?*

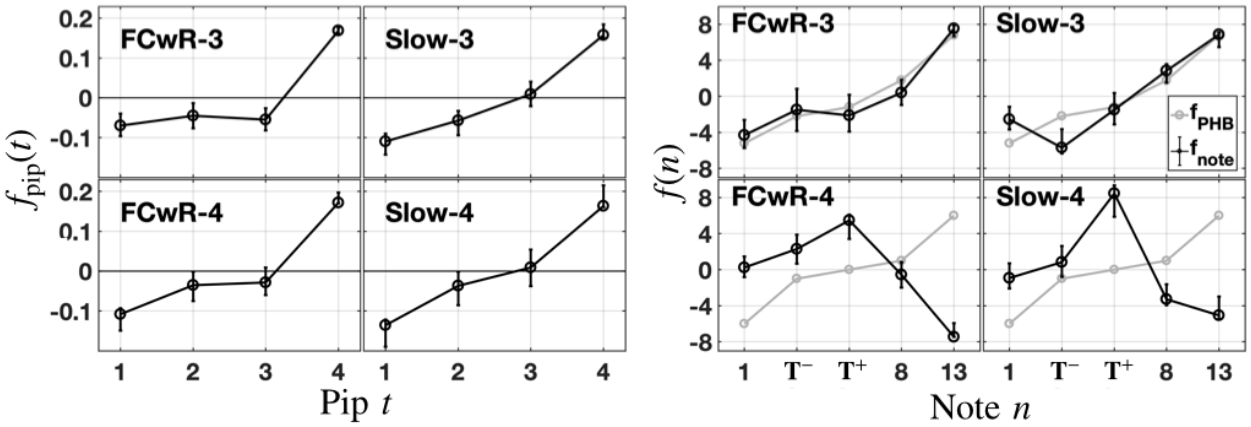

Figure 2.8: *The functions f_{pip} (left panel) and f_{note} (right panel) for the four tasks.* The gray lines in the right panel show the form of f_{note} predicted under the Pitch-height-biased model (Mednicoff et al. 2018). Markers show maximum likelihood estimates; error bars are 95% Bayesian credible intervals. Figure reproduced from [Ho, J., & Chubb, C. (2020). How rests and cyclic sequences influence performance in tone-scramble tasks. The Journal of the Acoustical Society of America, 147(6), 3859-3870.], with the permission of AIP Publishing.

2. *How do the biases depend on the temporal structure of the stimuli of a given task?*

Sequence-specific biases are influenced more strongly by the target notes of a stimulus than by temporal structure.

The stimuli in the FCwR-task variants differ strongly in temporal structure from the stimuli in the Slow-task variants. Individual tones last five times as long in the Slow-task variants than they do in the FCwR-task variants. In addition, the 4-note sequence that determines a stimulus in the Slow-task variants occurs only once, but is repeated five times in the FCwR-task variants.

Nonetheless, as revealed by the Note-function-biased model fits, the overall pattern of biases in the FCwR-3 (FCwR-4) task is similar to that in the Slow-3 (Slow-4) task. The left plot of Fig. 2.8 demonstrates that the last pip seems to play a more important role in the FCwR-task variants than in the Slow-task variants. For both FCwR-3 and FCwR-4 tasks, f_{pip} is fairly flat across pips 1, 2, and 3, and then jumps up abruptly on pip 4. By contrast, f_{pip} rises more gradually for the Slow-3 and Slow-4 tasks.

Differences between the note-functions of the 3-task variants and the note-functions of the

4-task variants are strikingly clear in the right panel of Fig. 2.8. The 3-task note-functions assign maximally positive values to note 13 (the high tonic) whereas the 4-task note-functions assign maximally negative values. In conjunction with the fact that the functions f_{pip} are maximally positive at pip 4, the note-functions for both 3-tasks imply that ending on note 13 biases listeners to respond "high-target" ("major"). In contrast, in both 4-tasks, ending on note 13 biases listeners to respond "low-target" ("fourth").

In itself, this observation might be taken to suggest that the 4-task note-functions are negatives of the corresponding 3-task note-functions. This would imply that the effects that operate in the 3-tasks to bias listeners to respond "major" also operate in the 4-tasks to bias listeners to respond "fourth." However, this does not appear to be true. Negating the note-functions for the FCwR-4 and Slow-4 tasks in the right panel of Fig. 2.8 does not convert them into the note functions of their 3-task counterparts. Furthermore, 4-task note-functions reach their maximal values at T^+ whereas both 3-task note-functions assign T^+ a value near 0. These findings imply that features of the stimuli other than the difference in pitch between the target notes are critical in determining the pattern of the sequence-specific biases in the task.

The difference in pitch between the target notes T^- and T^+ is the same in the FCwR-3, Slow-3, FCwR-4, and Slow-4 tasks; however, the sequence-specific biases are dramatically different between the two 3-task variants versus the two 4-task variants. This implies that it is not the difference in pitch between target notes that determines the biases. Plausibly, the pattern of sequence-specific biases in a given task is determined by the intervals formed between the two target notes and the context-defining notes 1, 8, and 13. Several features of the note-functions provide clues to the nature of this effect.

First, ending on note 13 (the high tonic) exerts a powerful influence on the biases in both 3- and 4-task variants; by contrast, ending on 1 (low tonic) exerts much less influence in the 3-task variants and little or no influence in the 4-task variants.

Music theory suggests that the high and low tonic should play similar roles in controlling

the scale-defined qualities of a tone sequence. The different roles of the high and low tonic in influencing the sequence-specific biases in both the 3- and 4-task variants thus suggest that the source of these biases may lie outside the scope of standard music theory. We discuss some possibilities in Sec. 2.7.

Sequence-specific biases may not be intrinsic to the system that is recruited to classify tone-scrambles.

If sequence-specific biases are intrinsic to tone-scramble classification, then we might expect these biases to operate with increasing strength in listeners with higher levels of scale-sensitivity. However, in each of the four tasks, the $f_\beta(R_k)$ are flat across listeners k with R_k above the median. In addition, in the Slow-3 and FCwR-4 tasks, $f_\beta(R_k)$ also appears to be greater than 0 for some listeners with R_k near or below 0.

2.7 General Discussion

Previous research (e.g., Chubb et al. (2013)) has shown that $\approx 70\%$ of listeners perceive little or no difference between major vs. minor tone-scrambles. Moreover, a single processing resource predominates in controlling performance in a range of tone-scramble tasks that use target notes unrelated to the difference between the major vs. minor scales (Dean & Chubb, 2017). This suggests that the resource recruited in these tasks confers general sensitivity to the qualities that music can achieve by establishing a tonic and selecting a scale, i.e., a distribution of intervals relative to the tonic used in the music. This led Dean and Chubb (2017) to call this resource "scale-sensitivity." Plausibly, the sensitivity of a listener to scale variations in actual music is also controlled (at least in part) by his-or-her level of scale-sensitivity.

The current study shows that the temporal structure of a task's stimuli exerts substantial influence on the ease with which listeners can extract scale-defined qualities from proto-

musical stimuli. For example, for $n = 3, 4$, scale-sensitivity facilitates performance with roughly twice the strength in the FRwR-n task vs. the Slow-n task.

Across the four stimulus temporal-structures tested (the FR-, FRwR-, FCwR-, and Slow- task variants) the facilitation strengths vary intuitively. For $n = 3, 4$, scale-sensitivity was found to facilitate performance in the FRwR-n task most strongly, less strongly and approximately equally in the FR-n and FCwR-n tasks, and most weakly in the Slow-n task.

Stimuli in the FRwR-n task include five bursts of four notes, with each burst containing a randomly-ordered sequence of the low tonic, target note, dominant, and high tonic (notes 1, **T**, 8, and 13, respectively). Stimuli in the FR-n task also contain five sets of these notes; however, they are presented in random order as a single, unbroken stream. Thus, the stimuli in the FRwR-n task work in two ways to structure the note sequence to enhance performance: (1) they break up the stream into separate bursts, and (2) they homogenize the stream by forcing each burst to contain one each of the four notes defining the stimulus. Either or both of these features may underlie the difference between $F_{\text{FRwR-}n}$ vs. $F_{\text{FR-}n}$ evident in Fig. 2.1.

Stimuli in the FCwR-n task have the same temporal structure as those in the FRwR-n task; however, each of the five bursts contains the same sequence of four notes (one each of notes 1, **T**, 8, and 13). As shown in Sec. 2.6, performance in the FCwR-n is undermined by sequence-specific biases. Plausibly, these biases tend to cancel out in the FRwR-n stimuli to yield the difference between $F_{\text{FRwR-}n}$ vs. $F_{\text{FCwR-}n}$ evident in Fig. 2.1.

The difference between $F_{\text{FCwR-}n}$ vs. $F_{\text{Slow-}n}$ is more interesting. Stimuli in the Slow-n task contain the same information as those in the FCwR-n task: in each case, the stimulus is defined by a 4-note sequence. Moreover, the total duration of the stimuli from both tasks is roughly equal. Nonetheless, $F_{\text{FCwR-}n}$ is substantially greater than $F_{\text{Slow-}n}$. This shows clearly that speeding up and repeating a sequence can increase the legibility of its scale-defined qualities. Increasing the frequency of occurrence of tones in a musical sequence is known to establish a stronger perception of tonal hierarchy (Knopoff & Hutchinson, 1983; C. Krumhansl, 1990; C. L. Krumhansl & Kessler, 1982; Youngblood, 1958; Rosenthal &

Hannon, 2016). However, increasing tone duration should enhance the perception of tonal hierarchy by an even greater magnitude (Lantz & Cuddy, 1998; Smith & Schmuckler, 2004). For example, (Lantz & Cuddy, 1998) found that when the total duration of tone sequences are held constant, the sequences that contain fewer tones of longer duration correspond to higher ratings of tonal stability. Our results and those of Mednicoff et al. (2018) regarding duration are at odds with these findings. The results of the current study may be explained by the repetition in the stimuli. Playing the same sequence of tones several times in a loop enhances the musicality of a tone sequence (Margulis & Simchy-Gross, 2016). Thus, perhaps the scale-defined qualities in our stimuli became clearer as a result of the increased musicality and the reorganization of the tones into a regularly-occurring rhythmic structure.

We might expect that using repetition in the Slow conditions (i.e., 5 repetitions of the same 4 tones) would improve performance in the Slow task; however, we are skeptical of this prediction. Based on anecdotal observations of our tasks, we predict that combining the slow speed of the tones with an overall longer stimulus length (resulting from the repetitions) would lead listeners to quickly become bored of the task and consequently not attend to the full stimulus length. Further, as demonstrated in Sec. 2.6, listeners are susceptible to sequence-specific biases for stimuli that are defined by a 4-note sequence.

The sequence-specific biases analyzed in Sec. 2.6 remain mysterious. This analysis reveals a striking difference in the pattern of biases in the 3-task variants vs. the 4-task variants. Notably, ending on note 13 (the high tonic) biases listeners to respond "high-target" ("major") in the FCwR-3 and Slow-3 tasks; by contrast, ending on the same note biases listeners to respond "low-target" ("fourth") in the FCwR-4 and Slow-4 tasks. The first part of this effect resembles an observation by Burnham, Long, and Zeide (2020) that listeners are more biased to categorize a melody as major (minor) if it is ascending (descending) in pitch. Burnham et al. (2020) speculate that listeners respond in this manner because major mode and ascending pitch both activate concepts related to positivity, while minor mode and descending pitch both activate concepts related to negativity. Mednicoff et al. (2018) suggest that the

bias in the 3-task may be explained by theories about the relationship between music and speech (Patel, 2005; Patel, Iversen, & Rosenberg, 2006). Specifically, intonation of speech that is spoken happily tends to end on a higher pitch (Juslin & Laukka, 2003; Swaminathan & Schellenberg, 2015; Curtis & Bharucha, 2010). Shared emotional expressiveness between music and speech would therefore suggest that a tone sequence that ends on a high note might be perceived as more happy (major). Although the happy-vs-sad distinction applies most naturally to major-vs-minor stimuli, we speculate that consonant-vs-dissonant stimuli might also differ along this spectrum. Listeners prefer consonance in music (Trainor, Tsang, & Cheung, 2002), which is associated with harmoniousness and stability based on pitch intervals (Meyer, 2008). On the other hand, the tritone interval (which is present in the high-target 4-task stimuli) is highly dissonant and associated with unpleasantness (Meyer, 2008; Plomp & Levelt, 1965), which listeners may relate to sadness more easily than the consonant fourth interval. Thus, if a tone sequence that ends on a high note is perceived as more happy, then perhaps listeners are biased to associate it with the more pleasant interval (fourth). Further research is needed to explore these ideas in depth.

Chapter 3

Mode Sensitivity Predicts Sensitivity to Speech Independently of Musical Training

3.1 Abstract

On each trial in the "3-task," the listener hears a rapid (923 BPM), randomly-ordered sequence comprising eight each of the notes in either an octave-doubled, G-major or G-minor triad and strives (with feedback) to judge which he-or-she heard. This task yields a bimodal distribution in performance with approximately 70% of listeners near chance and the other 30% near perfect. This study investigated whether performance in this task correlates with performance in processing speech prosody. Listeners were tested in the 3-task and also in two speech prosody tasks: In the Rating Task, listeners rated the similarity of pitch accent patterns between trisyllabic, nonsense words; in the Shape Task, listeners matched the pitch accent patterns of trisyllabic, nonsense words to visual pitch contours. Performance in the 3-task correlated significantly with performance in both prosody tasks even after effects due to years of musical training were removed. Conversely, years of musical training failed to predict

any additional variance in prosody-task performance after effects due to 3-task performance were removed. These results suggest that sensitivity to musical mode and speech prosody depend on shared processing resources that are largely immune to musical training.

3.2 Introduction

What is the nature of the relationship between music and speech perception? Music and speech are similar not only in their acoustic properties, but also in their use of pitch cues to convey information (e.g., Heffner and Slevc (2015); Juslin and Laukka (2003)). In music, for example, specific pitch intervals and contours can communicate mode, emotion, and boundaries of melodic phrases. Similarly, the rise and fall of a speaker's voice during speech can communicate emotion, lexical information in tonal languages, and boundaries of phrases or sentences. It is therefore unsurprising that substantial evidence supports an overlap in the processing of music and speech. For example, musical expertise correlates with skill in speech processing (Glushko, Steinhauer, DePriest, & Koelsch, 2016; Gordon, Magne, & Large, 2011; Hoch, Poulin-Charronnat, & Tillman, 2011; Ott, Langer, Oechslin, & Jancke, 2011; Roncaglia-Denissen, Bouwer, & Honing, 2018; Strait & Kraus, 2011; Patel, 2011). Musical training programs have been shown to improve speech-processing (Degé & Schwarzer, 2011; Kraus, Slater, Thompson, Hornickerl, & Strait, 2014). In children, phonological awareness is correlated with production and perception of pitch intervals (Loui, Kroog, Zuk, Winner, & Schlaug, 2011), and children with specific language impairment are also impaired relative to age-matched controls in various musical tasks using melodies as stimuli (Loui et al., 2011; Sallat & Jentschke, 2015). Furthermore, the ability to detect small frequency differences in musical stimuli seems to transfer to the ability to detect small frequency differences in linguistic stimuli, and this sensitivity improves with musical expertise (Schön, Magne, & Besson, 2004; Wong, Skoe, Russo, Dees, & Kraus, 2007).

The connection between music and speech likely has evolutionary roots. Juslin and

Laukka (2003) document many similarities in the emotional expressivity of music and speech and propose that sensitivity to the emotional content of both speech and music is conferred by a brain module that originally evolved to process non-verbal vocal expressions for their emotional content. Under this theory, the brain contains separate modules for processing speech and music; however, both modules make use of the evolutionarily more primordial module that confers sensitivity to the emotional meaning of non-verbal vocal expressions. Juslin and Laukka (2003) also argue that the similarities between speech and music in emotional expression only applies to properties that the performer can manipulate (e.g., tempo, loudness) and not to the more fixed properties in music such as harmony and mode. Thus, in particular, according to Juslin and Laukka (2003), the emotional content of a musical piece should not depend on its mode.

This conclusion is at odds with a substantial body of work supporting the claim that, on average across listeners, music in the major (Ionian) mode sounds "happy" whereas music in the minor (Aeolian) mode sounds "sad." (Cunningham & Sterling, 1988; Gagnon & Peretz, 2003; Gerardi & Gerken, 1995; Heinlein, 1928; Hevner, 1935; Kastner & Crowder, 1990; Leaver & Halpern, 2004; Peretz et al., 1998; Temperley & Tan, 2013). Because of this striking qualitative difference, the major and minor scales have come to play a central role in western music.

It has been argued that the "happiness" and "sadness" of music in the major and minor modes is in fact rooted in the resemblance of melodies in the major and minor modes to happy and sad speech respectively. For example, Huron (2008) begins with the observation that the prosodic variations in sad speech are suppressed in size compared to those in happy speech. He then analyzes a large corpus of minor- and major-key themes from Western classical music and shows that the average interval size was smaller for minor- than for major-key themes. Huron and Davis (2012) add to this story by showing that if one starts with melodies in the major mode and asks what modification of the major scale leads to melodies with the smallest variations in pitch, the answer is the harmonic minor scale. Cross-

cultural comparisons also support the idea that the association between musical modes and emotions is rooted in the vocal characteristics of different emotional states. For example, Bowling, Sundararajan, Han, and Purves (2012) show that (1) variations in emotional state alter the prosody of English and Tamil speech in similar ways, and (2) the tonal relationships used to express happiness and sadness in classical South Indian music parallel those used in Western music.

On the other hand, as Peretz (2002) notes, musical variation in pitch differs dramatically from prosodic variation in pitch. Nearly all music is constructed using a small, discrete set of pitches (the scale of the music). Typically one of the notes of the scale is singled out to serve as the "tonic" of the music. This note plays a special role in the music; for example, music typically starts and ends on the tonic, and ending on notes other than the tonic tends to make the music sound incomplete.

The current project addresses the following question: Is there overlap between the computational resources required for processing speech prosody and musical mode? On average across listeners, music in the major mode tends to sound "happy" whereas music in the minor mode tends to sound "sad" (Bonetti & Costa, 2019; Crowder, 1984, 1985a, 1985b; Cunningham & Sterling, 1988; Gagnon & Peretz, 2003; Gerardi & Gerken, 1995; Heinlein, 1928; Hevner, 1935; Kastner & Crowder, 1990; Leaver & Halpern, 2004; Peretz et al., 1998; Temperley & Tan, 2013). Further, Fourier spectra drawn from major music resemble those of happy or excited speech whereas spectra from minor music resemble those of sad or subdued speech (Juslin & Laukka, 2003; Curtis & Bharucha, 2010), suggesting that sensitivity to musical mode variations may be conferred by processes that originally evolved to extract emotional content from speech (Huron, 2008; Huron & Davis, 2012; Koelsch et al., 2004; Patel, 2005).

In the current study, we compared listeners' performance on the 3-task (measured as 3-task-d') to their accuracy on two speech prosody perception tasks. We expected that 3-task performance would positively correlate with prosody-task-performance. This result

alone, however, would not prove that sensitivity to speech prosody is heightened by mode sensitivity. An alternative possibility is that listeners who are highly sensitive to musical mode are more likely to seek out musical training, and musical training increases sensitivity to prosody. Under the latter story, years-musical-training should account for variance in prosody-task-performance beyond that accounted for by 3-task-d'. As we shall see, however, the reverse is true: 3-task-d' accounts for significant variance in prosody-task-performance that is not explained by years-musical-training; by contrast, years-musical-training fails to account for significant variance in prosody-task-performance that is not explained by 3-task-d'. This result suggests that mode sensitivity is important for processing prosody, and any apparent effect of musical training on prosody-task performance is due to the fact that listeners high in mode sensitivity are more likely to seek out musical training.

3.3 Methods

All methods were approved by the UCI Institutional Review Board.

3.3.1 Participants

Fifty-two listeners participated and were all undergraduates at the University of California, Irvine with self-reported normal hearing. All received course credit for their participation.

Twenty listeners reported having at least one year of formal musical training. To limit the influence of outliers, the number of years of musical training was counted as 12 if it was greater than 12 (N=2). With this restriction in force, the mean number of years of musical training across all listeners was 2.1 (standard deviation: 3.4).

Thirty-three listeners were native English speakers, and 15 listeners spoke only English. Thirty-seven listeners were multilingual. Eighteen listeners spoke Spanish, and 9 listeners reported that their primary language was tonal (Chinese dialects or Vietnamese). For these speakers, we unfortunately neglected to confirm that their primary language was also their

native language. Among the tonal language speakers, the mean number of years of musical training was 5.78 (standard deviation: 4.60). Among the non-tonal language speakers, the mean number of years of musical training was 1.86 (standard deviation: 4.40).

3.3.2 Procedure

The experiment consisted of three tasks, and task order was counterbalanced across listeners. Before beginning, listeners completed a general questionnaire to report their musical and language background. The only pieces of information from this questionnaire that are used below are the listener's (1) number of years of musical training, and (2) native language. The last 28 listeners to participate in the experiment also completed the Ollen Musical Sophistication Index (OMSI) questionnaire (Ollen, 2006), which yields a score that reflects the likelihood that a music expert would categorize the listener as "more musically sophisticated" (e.g., knowledgeable about music, can play an instrument or sing, can understand and create music).

The experiment took place in a quiet lab on a Windows Dell computer with a standard Realtek audio/sound card using Matlab. Stimuli in the 3-task were presented at a rate of 50000 samples/s. Listeners wore JBL Elite 300 noise-cancelling headphones with volume adjusted to their comfort level.

3.3.3 The 3-task

Stimuli were tone-scrambles, which are sequences of 32 randomly-ordered pure tones comprising 8 copies each of the following notes from the standard equal-tempered chromatic scale: G_5 (783.99 Hz), D_6 (1174.66 Hz), and G_6 (1567.98 Hz). Major stimuli also contained 8 copies of B_5 (987.77 Hz); minor stimuli contained 8 copies of $B\flat_5$ (932.33 Hz). The duration of each tone was 65 milliseconds, for a stimulus duration of approximately 2.08 seconds. Each individual tone was windowed by a raised cosine function with a 22.5-ms rise time.

Before beginning the task, listeners heard eight, visually labeled, example stimuli that al-

ternated between "Type 1 (Major/Happy)" and "Type 2 (Minor/Sad)." Then, on each trial, listeners heard a single stimulus and were asked to classify it as "Type 1 (Major/Happy)" or 'Type 2 (Minor/Sad)." Correctness feedback was printed to the screen after each trial, and proportion correct was given at the end of each block. Listeners completed two blocks of 50 trials.

3.3.4 Speech shape task

Stimuli

We used 36 recordings of trisyllabic, nonsense words that were completely voiced and contained a pitch accent on the first, second, or third syllable. These recordings were drawn from a stimulus database developed by Gupta et al. (2004). The words in this database, which are all spoken by a female native speaker of American English, contain CV non-final syllables and a CVC final syllable (e.g., ba-le-vel). We were able to obtain trisyllabic words from the database that contained pitch accents on the first or second syllable, but not the third syllable. Therefore, we modified some stimuli to create words with a pitch accent on the third syllable. To make this set of words, we selected five-syllable words from the database (which contained a pitch accent on the third syllable) and smoothly edited out the fourth and fifth syllables, resulting in trisyllabic words with a pitch accent on the third syllable.

Visual pitch contour curves were generated for each word using Praat software (Boersma & Weenink, 2017). To create these shapes, the pitch contour was extracted from each stimulus recording via autocorrelation and smoothed (see Figure 3.1 for example contours). The stimulus contour images are available for download at https://github.com/joselyngithubs/Speech-Stimuli/.

Task

At the start of the task, listeners heard 9 example words and viewed their corresponding visual pitch contour shapes. They could play these examples repeatedly. When they were

ready to begin, they completed 36 trials of the task. On each trial, they listened to a word and viewed three pitch contour shapes. They were instructed to select the shape that best matched the pitch contour of the word that they had heard (see Figure 3.1 for a sample screen display). They had the option to replay the word before selecting the shape. They received feedback on each trial that indicated which of the three options was correct, and they received proportion correct at the end of the task.

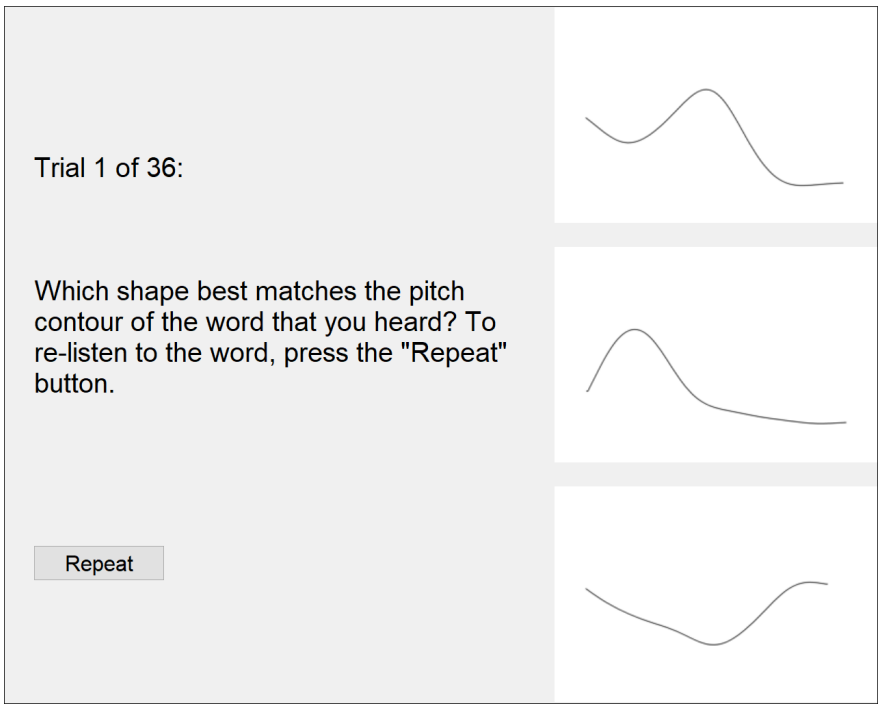

Figure 3.1: A sample trial of the Speech Shape Task. The listener heard a word and selected the pitch contour shape that best matched the word's pitch pattern. The listener had the option to replay the word before making their selection.

3.3.5 Speech rating task

Stimuli

Seventy-two recordings of trisyllabic, nonsense words were selected from the stimulus database by Gupta et al. (2004). Some of these stimuli were identical to those of the Speech shape Task. Each word contained a pitch accent on the first, second, or third syllable.

Words were presented as 36 pairs, and all listeners received the same pairs in different

order. Pairs were arranged such that the possible combinations of pitch accent positions in the two words were counterbalanced as fully as possible. Each possible word occurred only once in the entire stimulus set. Half of the word pairs had matching accent positions.

Task

Before beginning the task, listeners heard three example trials of two words with matching pitch accent positions, and three example trials of two words with different pitch accent positions. They could listen to the examples repeatedly. When they were ready, they completed 36 trials of the task. On each trial, they listened to two words (played sequentially) and clicked a button on the screen to rate the similarity of the two words' pitch accent positions on a scale of 1-6 (1 = extremely dissimilar, 6 = extremely similar). They had the option to re-listen to the words before they selected their rating.

3.4 Results

Listeners' performance in the 3-task is summarized by d' values, using the last 75 trials of the task (the first 25 trials were treated as practice). If a listener responded correctly to all n of the major (minor) stimuli in those last 75 trials, then the probability of a correct response was adjusted to $\frac{n-0.5}{n}$ (as suggested by Macmillan and Kaplan (1985)). This implies that a d' value around 4.4 corresponds to near-perfect performance on the 75 trials. (A d' value of 0 corresponds to chance performance.)

Fig. 3.2 displays the histogram 3-task-d' values. As found in other studies, the distribution of 3-task-d' values has a large concentration of listeners near 0 and is skewed positive. In the current sample, 64% of the listeners had 3-task-d' values less than 1 (which corresponds to proportion correct ≤ 0.69).

As seen in Fig 3.3, listeners' 3-task-d' values correlated with their self-reported years of musical training ($r = 0.54$, $p = 0.000$). However, as seen in Fig. 3.2, this correlation is

driven primarily by a large group of low-performers with no musical training.

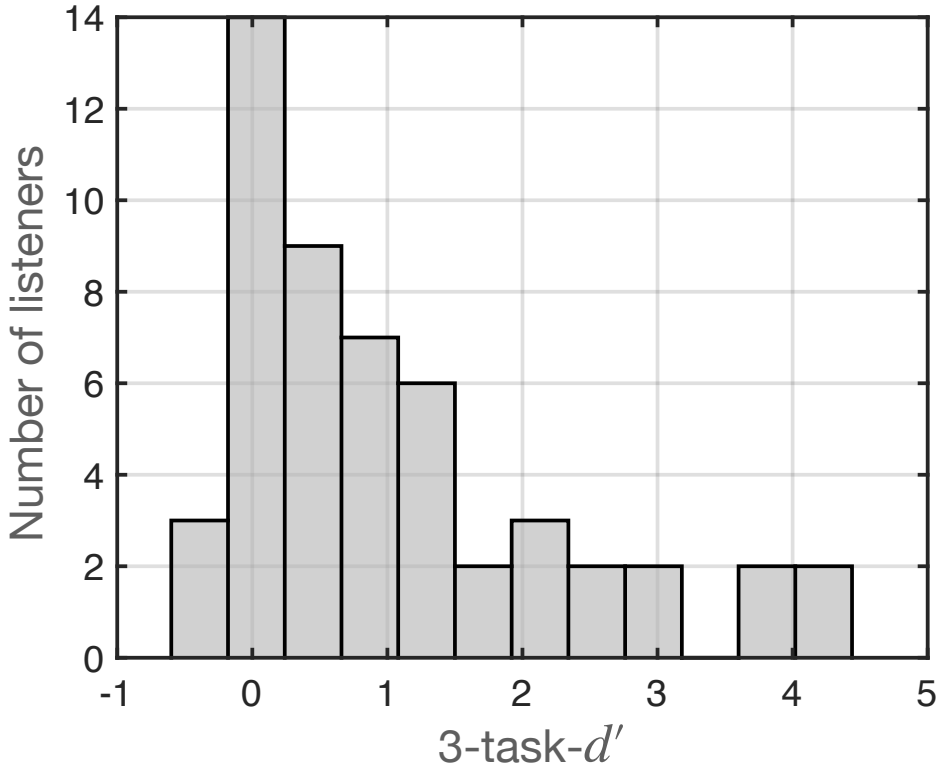

Figure 3.2: Histogram of 3-task-d' values (estimated from the final 75 out of 100 trials) achieved by the 52 listeners.

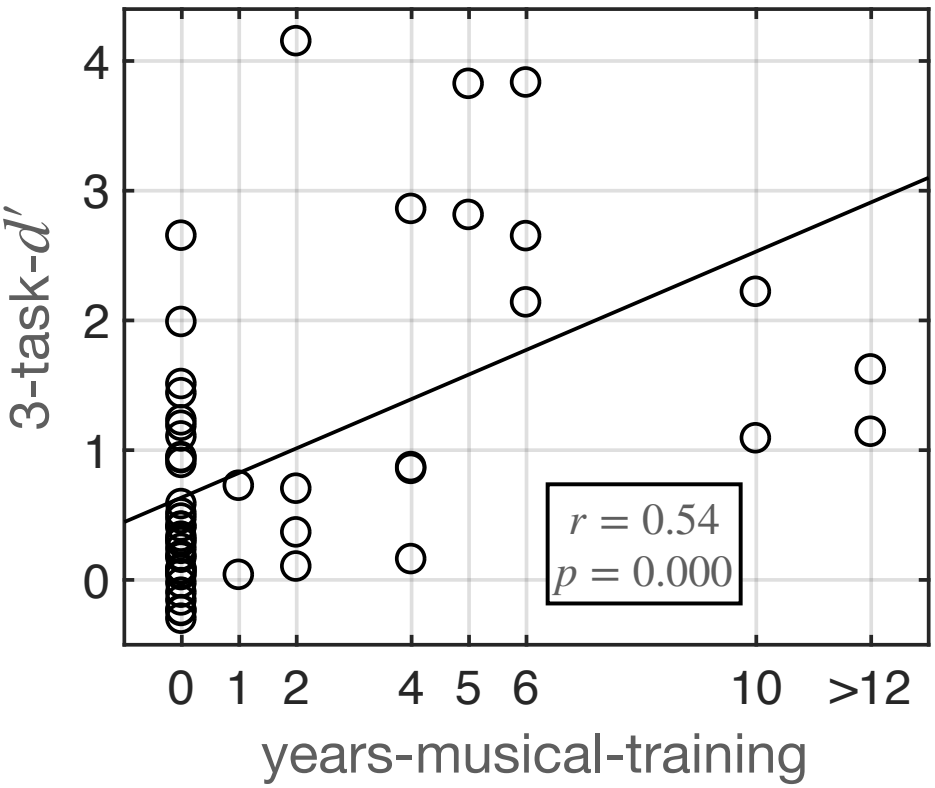

Figure 3.3: The relationship between musical training and 3-task-d'.

The last 28 listeners who participated in the experiment also took the Ollen Musical Sophistication Index questionnaire. Their 3-task-d' values did not correlate significantly with their musical sophistication index ($r = 0.08$, $p = 0.69$). However, none of these listeners achieved an index above 500, which is the score corresponding to "more musically sophisticated," and most of these listeners only had 0-2 years of musical training. Thus, this sample does not provide a strong test of the relationship between Ollen score and performance in the 3-task.

For each participant in each of the two prosody tasks, a score was derived by applying the logit function to proportion correct. The scores for the two prosody tasks were strongly correlated ($r = 0.91$, $p = 0.000$). Therefore, we took the average of the scores for the two tasks to compute an overall prosody-score for each listener.

Fig 3.4A shows that across our 52 listeners, 3-task-d' was positively correlated with prosody-score ($r = 0.68$, $p = 0.000$). Similarly, Fig 3.4B, shows that years-musical-training

was correlated with prosody-score ($r = 0.42$, $p = 0.002$). Fig. 3.4C shows that when we orthogonalize the 52-dimensional vector of 3-task-d' values with respect to the corresponding 52-dimensional vector of years-musical-training, the resulting vector still has a highly significant, positive correlation with prosody-score; by contrast, Fig. 3.4D shows that when we orthogonalize years-musical-training with respect to 3-task-d', the resulting vector does not correlate significantly with prosody-score. This shows that 3-task-d' overshadows years-musical-training as a predictor of prosody-score in the following sense: 3-task-d' explains significant variance in prosody-score that cannot be explained by years-musical-training; however, years-musical-training fails to explain any appreciable variance in prosody-score that cannot be explained by 3-task-d'. Thus, the results shown in Fig. 3.4 suggest that the correlation between years-musical-training and prosody-score is produced indirectly by (1) the relation between years-musical-training and 3-task-d' and (2) the relation between 3-task-d' and prosody-score.

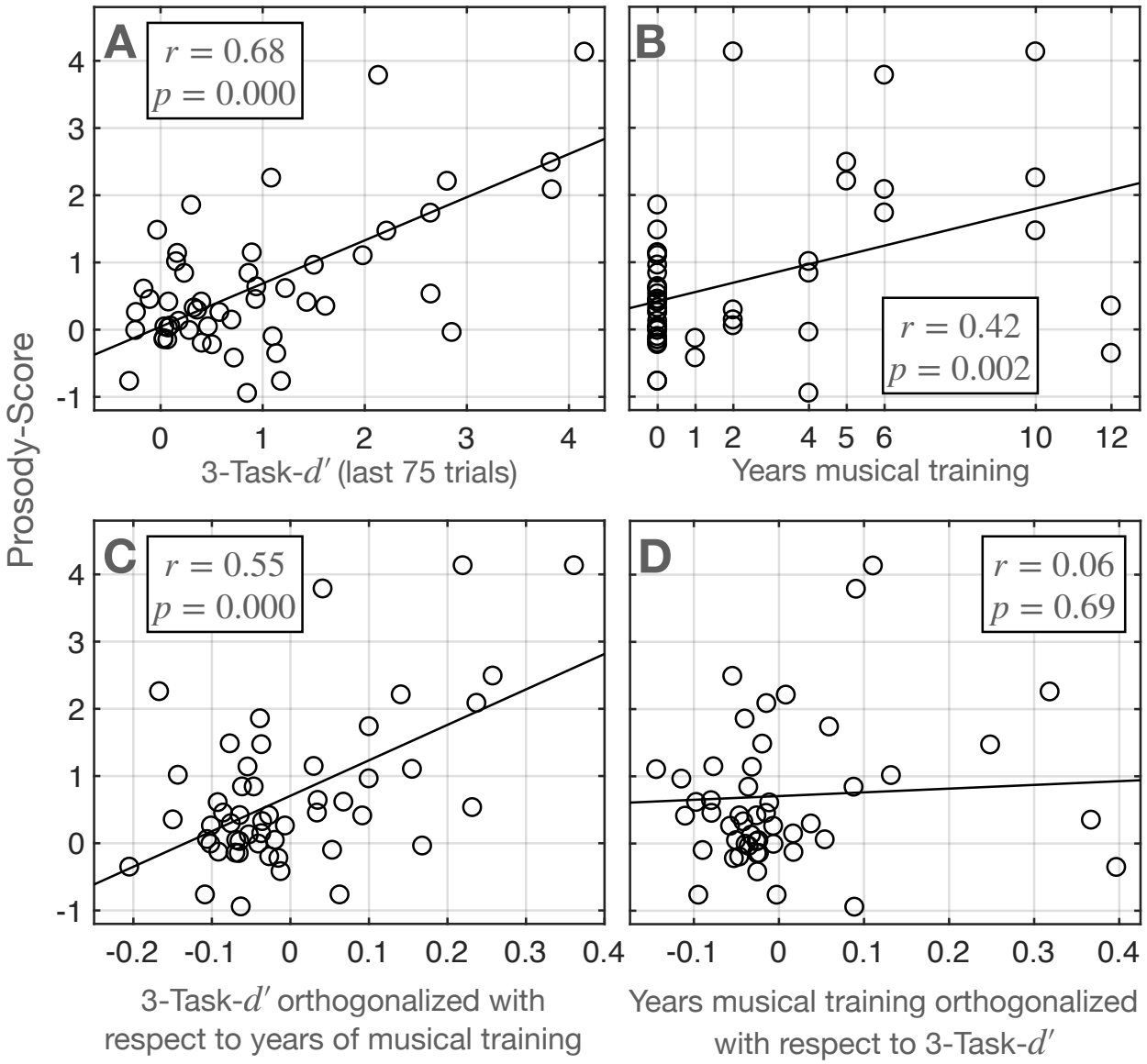

Figure 3.4: Listeners' prosody-task summary scores plotted against (A) 3-task-d', (B) Years-of-musical-training, (C) 3-task-d' orthogonalized with respect to Years-of-musical-training, (D) Years-of-musical-training orthogonalized with respect to 3-task-d'. The prosody-task summary scores reflect the average of (1) the logit function applied to proportion correct in the rating task and (2) the logit function applied to proportion correct in the shape task.

3.5 Discussion

The current experiment shows that 3-task-d' is correlated with prosody-score. This suggests that some underlying factor, rooted in either or both of the genetic make-up and/or expe-

rience of our listeners, influences 3-task-d' and prosody-score in a similar way. Our analysis indicates that this underlying factor is not musical training. All effects of years-musical-training on prosody-score seem to be mediated by 3-task-d'.

What then might this underlying factor be? We shall argue that it is sensitivity to differences in pitch across time. Evidence supporting this contention comes from Mann (2014) who tested 111 listeners in the 3-task as well as several other basic musical tasks, one of which, the "pitch comparison task," is especially relevant for current purposes. Each listener performed 200 trials in the 3-task. The values of 3-task-d' plotted in Fig. 3.5 are derived from the last 150 trials. On each trial in the pitch comparison task, the listener heard two, 500 ms pure tones (windowed by a raised cosine with a 20 ms rise-time) separated by 1-sec of silence and had to judge whether the second tone was higher or lower in pitch than the first. The frequency f_1 of the first tone was random, uniformly distributed on the linear frequency interval from 300 to 2000 Hz. The frequency f_2 of the second tone was equally likely to be higher or lower than f_1, and the difference between f_2 to f_1 was adaptively controlled by two, randomly interleaved staircases. Correctness feedback was given after each trial. Each listener performed 100 trials in this task, and the pitch difference threshold (i.e., the absolute value of the log of the frequency ratio, expressed in cents in Fig. 3.5) was estimated for each listener. Specifically, a Weibull function was fit to the data for each listener, and the threshold was taken to be the absolute number of cents between f_1 and f_2 required for the listener to achieve 80% correct.

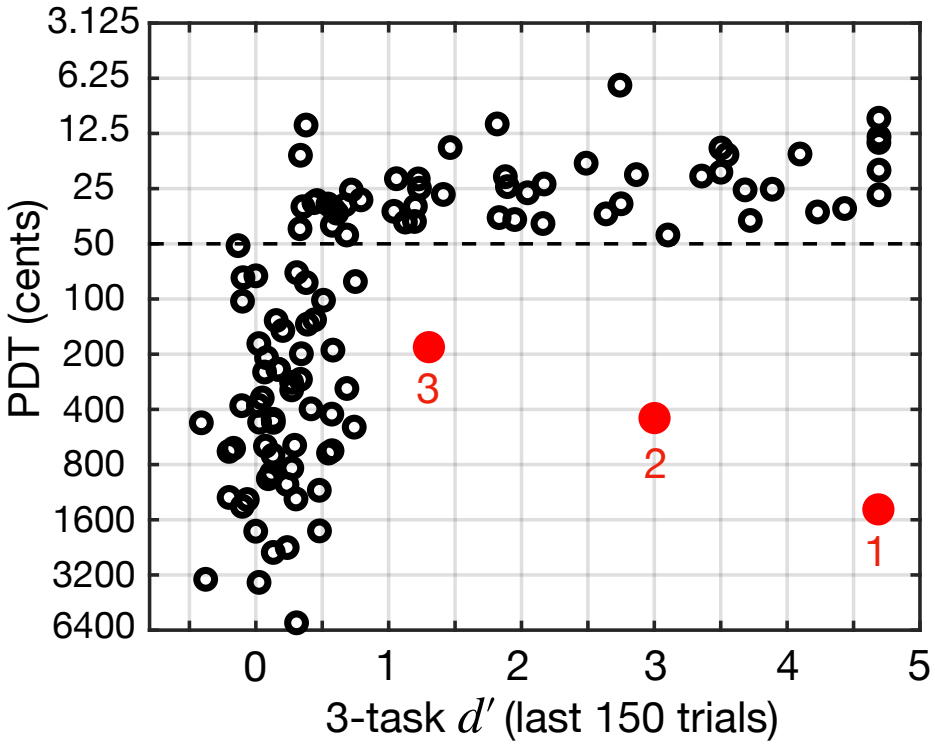

Figure 3.5: Scatterplot of pitch-difference-threshold (PDT) as a function of 3-task-d' (from Mann, 2014). PDT's are plotted (on a log scale) with values decreasing from bottom to top to reflect increasingly good performance. The dashed line is at 50 cents (a quarter-tone). Out of the 59 listeners whose PDT's were higher than 50 cents, only 3 listeners (the filled circles) achieved d' values greater than 0.75 (which corresponds to proportion correct ≤ 0.65) in the 3-task.

The scatter plot of the results from Mann (2014) of 3-task-d' vs. pitch-difference-threshold (PDT) is shown in Fig. 3.5. In this figure, PDT's are plotted on a log scale decreasing from bottom to top to reflect increasing levels of performance. There are two things to note about this figure:

1. All except three of the 59 listeners whose PDT's were higher than 50 cents performed near chance in the 3-task. The 3 listeners marked with filled circles are outliers – these listeners most likely did not give the same amount of effort in all tasks.

2. Across the 52 listeners whose PDT's are lower than 50 cents, the distribution of 3-task-d' values is approximately uniform from 0 to ceiling.

It is striking that (excluding the 3 outliers) the PDT-value required for high performance

in the 3-task is 50 cents (a quarter tone)–i.e., the interval exactly half way between two successive notes in the chromatic scale, and in particular between the two notes, $B\flat_5$ and B_5 that differ in the major and minor tone-scrambles in the 3-task. The relationship between PDT and 3-task d' reveals that a threshold less than 50 cents is necessary to successfully differentiate the major-vs-minor stimuli of the 3-task. However, simply having this low threshold is by no means sufficient to guarantee good performance in the 3-task.

Overall, these results from Mann (2014) strongly suggest that listeners with high PDT are impaired in comparing two distinct pitches across time. To consider the results of the current experiment in this context, we note that 82% of the listeners with 3-task-$d' < 0.75$ in Fig. 3.5 had elevated PDT. There are 26 listeners in the current experiment with 3-task-$d' < 0.75$. The results of Mann (2014) might suggest that of these 26 listeners, roughly $21 = 0.82 \times 26$ have elevated PDT. The mean prosody score across the 26 listeners with 3-task-$d' < 0.75$ is 0.44; by contrast, the mean prosody score across the other 26 listeners with 3-task-$d' > 0.75$ is 1.17. This difference suggests that just as 3-task-d' is suppressed for listeners with elevated PDT, so is prosody-score.

A follow-up experiment (data collection still ongoing) supports this prediction. 93 listeners were tested in a series of tasks including the Shape Task and a pitch comparison task that was similar to the task used by Mann (2014). The left panel of Fig. 3.6 shows the scatter plot of 3-task-d' vs. pitch-difference-threshold (PDT). The right panel of Fig. 3.6 shows the scatter plot of the logit of the Shape Task scores vs. PDT. Across the 46 listeners with PDT < 50, Shape Task score correlates with 3-task-d' by $r = 0.29, p < 0.05$. Across the 47 listeners with PDT > 50, Shape Task score correlates with 3-task-d' by $r = 0.23, p = 0.12$. These results suggest that pitch-difference-threshold, rather than musical training, exerts strong and similar influences on both the sensitivity to musical mode and the sensitivity to pitch prosody of speech. We continue to explore the relationship between 3-task-d' and pitch-difference-threshold in the next chapter.

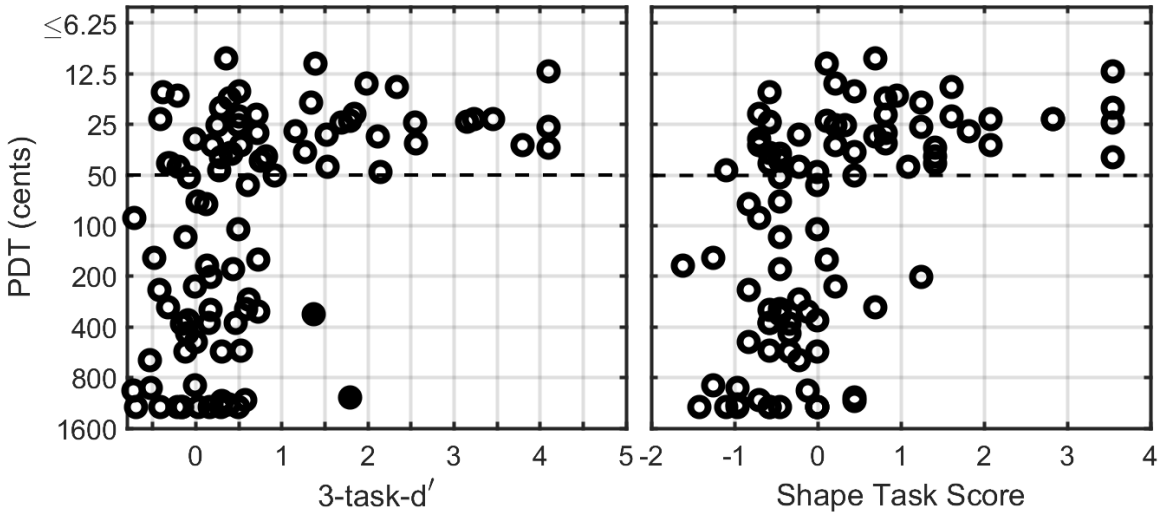

Figure 3.6: Scatterplots of 3-task-d' vs. PDT (left) and Shape Task score vs. PDT (right). PDT's are plotted (on a log scale) with values decreasing from bottom to top to reflect increasingly good performance. The dashed line is at 50 cents (a quarter-tone). Dark circles represent the 2 listeners whose PDT's were higher than 50 cents but achieved d' values greater than 0.75 (which corresponds to proportion correct ≤ 0.65) in the 3-task.

A large body of research shows that musical training is associated with a wide range of heightened abilities. For example, musicians are better than non-musicians at discriminating simple tones (Buss, Taylor, & Leibold, 2014; Fujioka, Trainor, Ross, Kakigi, & Pantev, 2004, 2005; Micheyl, Delhommeau, Perrot, & Oxenham, 2006) and complex melodic stimuli (Pantev et al., 1998). They also perform better than non-musicians in tasks requiring sound segregation (Parbery-Clark, Skoe, Lam, & Kraus, 2009), auditory attention (Strait, Kraus, Parbery-Clark, & Ashley, 2010), speech-processing (Besson, Chobert, & Marie, 2011a, 2011b; Marie, Magne, & Besson, 2010; Marie, Delogu, Lampis, Belardinelli, & Besson, 2011; Morrill, Devin, Dilley, & Hambrick, 2015; Parbery-Clark, Strait, Anderson, Hittner, & Kraus, 2011) as well as executive control (Bialystok & DePape, 2009; Zuk, Benjamin, Kenyon, & Gaab, 2014). The results of the current experiment illustrate the hazards in trying to use correlation to show that musical training heightens some particular skill. When one observes a positive correlation between musical training and performance in a given task, it seems natural to conclude that musical training is the cause of the improvement. However, this need not

be true. For example, our findings suggest that (despite the positive correlation between years-musical-training and prosody-score) musical training may well have little or no effect on prosody-score. To recapitulate the reasoning behind this claim:

1. Musical training seems to have little or no effect on scale-sensitivity. The positive correlation between years-musical-training and 3-task-d' is plausibly due to the fact that listeners with high scale-sensitivity are more likely to seek out musical training than listeners with low scale-sensitivity. This conclusion is bolstered by the finding that 6-month-old infants show the same distribution of sensitivity in the 3-task as adults (Adler et al., 2020).

2. Scale-sensitivity explains nearly all of the variance in prosody-score that is explained by years-musical-training; however, scale-sensitivity also explains a large (highly significant) proportion of variance in prosody-score that years-musical-training fails to explain. Thus 3-task-d' almost entirely overshadows years-musical-training as a predictor of prosody-score. This is exactly what one would expect if musical training in itself has no effect on prosody-score.

These observations echo those of Correia et al. (2020) who showed that skill in recognizing vocal emotions is better explained by innate musical predispositions than by musical training. These authors express well-founded skepticism that musical training can improve performance on tasks that are unrelated to music.

We propose that in any study investigating the relationship between musical training and performance in some target task, the scale-sensitivity (SS, which can be represented by 3-task-d') of all listeners should be measured as a matter of standard practice. The reasons are as follows:

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 3-task and estimating d' from the last 50 trials.

4. Only by partialling out effects on target task performance due to SS can one determine whether musical training accounts for any of the variance in performance beyond the effects due to SS.

Overall, our results support a connection between music and speech. However, the nature of the link between prosodic sensitivity and musical sensitivity remains unclear. Patel (2011) proposed that musical sensitivity drives speech sensitivity, which would explain why musicians might experience music-to-language benefits. Musical processing is more demanding than speech processing due to its precision (e.g., listeners are sensitive to fine-grained changes in music, whereas speech can still be understood under broad changes); further, musical experiences are emotionally rewarding. Therefore, this hypothesis suggests that musical processing is more likely to promote plasticity in the networks shared with speech processing. Abundant research suggests that the reverse is also possible. For example, experience with a tonal language seems to confer more precise pitch processing of music and speech (Gandour et al., 2000; Krishnan, Xu, Gandour, & Cariani, 2005; Deutsch, Henthorn, Marvin, & Xu, 2006). Therefore, the influence between the music and language domains may be bidirectional (Bidelman, Hutka, & Moreno, 2013; Asaridou & McQueen, 2013).

Although we can not directly resolve these theories with our results, we can provide a case that emotion is the common agent that links music and speech prosody. Music shares many parallels with emotional speech, predominantly in pitch; for example, the spectra of major and minor music resemble those of happy and sad speech, respectively (Juslin & Laukka, 2003; Curtis & Bharucha, 2010). Musicianship has been observed to correlate with recogni-

tion of emotions in speech prosody (Farmer, Jicol, & Petrini, 2020; Lima & Castro, 2011). Evidence also suggests that greater emotional intelligence, rather than musical expertise, corresponds with more accurate recognition of emotions in speech and music (Trimmer & Cuddy, 2008). In the context of our study, we speculate that major-vs-minor (happy vs. sad) tone-scramble discrimination ability extends to speech prosody perception to aid the processing of emotional content in speech. Given the bimodal distribution of performance on the 3-task, our hypothesis would imply that only a small sample of the population are highly sensitive to the emotional content in speech and music, while the majority of people are not. Further research is needed to assess this possibility, since we did not directly assess emotional processing in the current study.

We note some limitations in the study. First, the pitch and intensity contours of the stimuli in the Speech Shape Task are correlated ($r = 0.67$, $p = 0.00$). Although listeners were instructed to only pay attention to the words' pitch patterns, it is possible that the listeners made their judgments based on each word's intensity pattern instead of (or in addition to) the pitch pattern. Second, we did not assess whether our sample of listeners have amusia (tone deafness). However, we administered a pitch discrimination task to the last 28 listeners and found that no listener scored below 70% accuracy. We believe this provides sufficient support that low performance on the tasks is not a result of tone deafness.

3.6 Conclusion

In summary, we found that sensitivity to mode (major-vs-minor) is related to sensitivity to pitch variations in speech prosody. Additionally, any apparent influence of musical training on sensitivity to prosody is actually due to differences in sensitivity to mode. These findings suggest that sensitivity to variations in mode and sensitivity to variations in prosody depend in part on overlapping mechanisms. We speculate that these mechanisms support the processing of emotional content.

Chapter 4

Many listeners have roved pitch-comparison thresholds above a quarter-tone; very few can discriminate major from minor tone-scrambles

4.1 Abstract

On each trial in the "3-task," the listener hears a rapid, random sequence of tones containing equal numbers of notes of either a G-major or G-minor triad and strives (with feedback) to judge which type of "tone-scramble" they heard. This task yields a dramatically bimodal distribution of performance. On each trial in a "pitch-comparison task," the listener hears two tones and judges whether the second tone is higher or lower than the first. When the first tone is roved (rather than fixed throughout the task), performance varies dramatically across listeners, with median threshold $\approx$ 50 cents. Strikingly, nearly all listeners with

thresholds higher than 50 cents performed near chance in the 3-task. Across listeners with thresholds below 50 cents, 3-task performance was uniformly distributed from chance to ceiling; thus, the large, lower mode of the distribution in 3-task performance is produced mainly by listeners with roved pitch-comparison thresholds greater than 50 cents.

4.2 Introduction

Western theories of music composition universally agree that variations in musical scale are central to the meaning that music can convey (Rameau, 1971–orig., 1722; Schoenberg, 1978–orig. 1922; Tymoczko, 2011), and within this tradition, the difference between the major and minor scales is fundamental. At the core of this difference is the triad composed of the tonic (scale degree 1), the dominant (scale degree five) and the mediant (scale degree three). Among all seven scale degrees, the mediant is unique in the following respect: it alone differs in the major scale and in all common variants of the minor scale. In the major scale, the mediant is four semitones above the tonic, and in all of the natural, harmonic, ascending melodic, and descending melodic minor scales the mediant is a semitone lower.

Thus, one might expect the qualitative difference between the major and minor scales to be very vividly expressed by the major and minor stimuli used in the "3-task" (Chubb et al., 2013; Dean & Chubb, 2017; Mednicoff et al., 2018; Ho & Chubb, 2020; Adler et al., 2020). However, the 3-task yields a dramatic, bimodal distribution in performance: approximately 70% of listeners perform near chance while the remaining 30% perform near ceiling. This result is consistent with previous findings of Blechner (1977) and Crowder (1985b) who observed a similar bimodal distribution of performance in tasks requiring listeners to classify triadic chords as major vs. minor. Previous research has ruled out musical training as the source of high performance on the 3-task (Chubb et al., 2013; Dean & Chubb, 2017; Mednicoff et al., 2018; Ho & Chubb, 2020).

This study investigates the possibility that individual differences in basic pitch-processing

ability play a role in producing the bimodal distribution in 3-task performance. In particular, we focus our attention on the relationship between performance in the 3-task and performance in "roved pitch-difference" tasks (RPD-tasks). In an RPD-task, the listener hears two pure tones on each trial; the first is chosen randomly from a large range of frequencies, and the task is to judge whether the second tone is higher or lower than the first.

Building on previous studies focused on listeners with cortical lesions (Johnsrude, Penhune, & Zatorre, 2000; Tramo, Shah, & Braida, 2002), Semal and Demany (2006) showed that there exist listeners with otherwise normal hearing for whom RPD-tasks are highly challenging for the following, unexpected reason: although they can tell when the two tones in a given trial are different, these listeners are markedly impaired at discerning the direction of the difference. In the main experiment of Semal and Demany (2006), the listener heard two pairs of pure tones on each trial. In one pair, the tones were identical; in the other pair, the tones differed in frequency. In the "detection" task, the listener judged which tone-pair contained the change (without reporting the direction of the change). In the "identification" task, the listener judged the direction of the change (without reporting which pair contained the change). Semal and Demany (2006) demonstrated that for some listeners (whose hearing was otherwise normal), the threshold frequency difference for the identification task was substantially higher than the threshold difference for the detection task. We shall call such listeners "RPD-challenged." Mathias, Micheyl, and Bailey (2010) replicated the experiment of Semal and Demany (2006) and showed in addition that the difficulties experienced by RPD-challenged listeners (1) are greatly decreased if the first tone is fixed across trials (i.e., if the rove is removed), and (2) are most dramatic when the first tone is roved across a very wide range of frequencies.

Each of the studies of Semal and Demany (2006) and Mathias et al. (2010) included unimpaired listeners as well as RPD-challenged listeners. The RPD-challenged listeners used in both studies were selected using a screening process in which a large group of potential listeners were pretested (in the same room) in restricted versions of the identification and

detection tasks. Only listeners whose performance was substantially worse on the identification test compared to the detection test were used in the actual experiment. We will refer to the threshold frequency difference in the RPD-task as the listener's PDT (for "Pitch Difference Threshold").

A study by Mann (2014) looked at how RPD-task performance is distributed across a sizeable group of listeners. On each trial in this study's pitch difference task, the listener heard two, 500 ms pure tones (windowed by a raised cosine with a 20 ms rise-time) separated by 1-sec of silence and had to judge whether the second tone was higher or lower in pitch than the first. The frequency f_1 of the first tone was random, uniformly distributed on the linear frequency interval from 300 to 2000 Hz. The frequency f_2 of the second tone was equally likely to be higher or lower than f_1, and the difference between f_2 and f_1 was adaptively controlled by two, randomly interleaved staircases. Correctness feedback was given after each trial. Each listener performed 100 trials in this task, and the threshold pitch difference (i.e., the absolute value of the log of the frequency ratio, expressed in cents in Fig. 3.5) was estimated for each listener. Specifically, a Weibull function was fit to the data for each listener, and the threshold was taken to be the absolute number of cents between f_1 and f_2 required for the listener to achieve 80% correct. The results for this task revealed that PDT's vary widely across listeners, ranging from around 6 cents to around 2.5 octaves (3000 cents). Slightly more than half (59) of the 111 listeners tested had PDT's greater than 50 cents (1 quarter-tone). For this low-performing subset of listeners, PDT's occur with probability that is approximately inversely proportional to their magnitude; thus, there are roughly equal numbers of PDT's in each of the following intervals: 50-to-100 cents, 100-to-200 cents, 200-to-400 cents, 400-to-800 cents, 800-to-1600 cents, 1600-to-3200 cents (Fig. 3.5).

In the study by Mann (2014), these 111 listeners were also tested in the 3-task. Strikingly, as shown in Fig. 3.5, nearly all of the listeners with PDT's greater than a quarter-tone perform very near chance in the 3-task. By contrast, for listeners with PDT's lower than a quarter-tone, sensitivity in the 3-task (as reflected by d') is uniformly distributed from

chance to ceiling.

The dark bars of the upper (lower) panel in Fig. 4.1 plot the histogram of d' (proportion correct) in the 3-task across all 111 listeners. Consistent with previous studies (Chubb et al., 2013; Dean & Chubb, 2017; Mednicoff et al., 2018; Ho & Chubb, 2020), the histogram of 3-task-d' has a large mode near 0 and strong positive skew, and the histogram of proportion correct is bimodal with one mode near 0.5 (chance performance) and another at 1.0 (ceiling). The lighter bars in the upper (lower) panel of Fig. 4.1 show the distribution of d' (proportion correct) in the 3-task when the listeners who achieved PDT's above 50 cents are excluded. The large peak near chance performance ($d' = 0$ and proportion correct $= 0.5$) is greatly reduced in each panel of Fig. 4.1 resulting in a roughly uniform distribution of 3-task-d' and a distribution of proportion correct with a single prominent mode at ceiling performance.

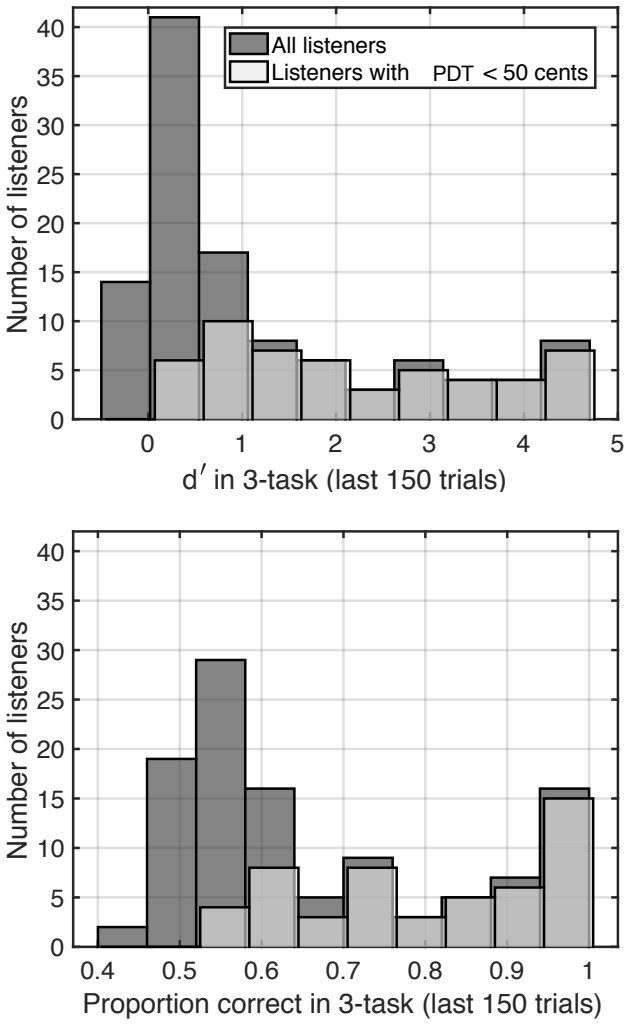

Figure 4.1: Histograms of d' values in the 3-task (across the last 150 of 200 trials) from Mann (2014). Dark gray bars show the histogram for all listeners. The light gray bars show the histogram for only those listeners who achieved PDT's lower than 50 cents.

We can conclude the following points based on these results:

1. Having a PDT less than a quarter-tone is an important precondition to perform well in the 3-task; however, it is not sufficient to insure high performance: there exist many listeners with PDT's below a quarter-tone who nonetheless perform poorly in the 3-task.

2. The listeners with PDT's greater than a quarter-tone produce the lower mode in the bimodal distribution in 3-task performance; when they are removed from the sample of

listeners, the distribution of 3-task performance becomes uniform. If proportion correct in the 3-task is used as the measure of performance (instead of d'), this distribution becomes unimodal (with the mode at ceiling).

The current experiment explores how variations in the pitch-comparison task influence this pattern. A new group of listeners is tested in the 3-task as well as in four pitch-comparison tasks. Previous research suggests that fixing the first tone across trials in a pitch-comparison task makes the task much easier for nearly all listeners (e.g., Mathias et al. (2010)). A possible reason for the improved performance observed in such "fixed pitch-comparison" (FPC) tasks is that fixing f_1 enables the listener to create a durable, internal representation of f_1 across trials with which to compare f_2. Such a strategy is not available in RPD-tasks. This suggests that low performers in RPD-tasks may have difficulty preserving a temporary memory of f_1 for comparison with f_2. If so, then perhaps RPD-task performance will degrade for RPD-challenged listeners if the delay between tone-1 and tone-2 is increased. To investigate this question, we include two RPD-tasks, one with inter-tone interval (ITI) 0.5 sec. and the other with ITI 1.0 sec.

Finally, we include a task in which f_2 can be either higher, lower or equal to f_1, and the listener is tasked with classifying the stimulus accordingly. If a listener can hear that f_2 differs from f_1 but cannot discern the direction of the difference, then on trials in which f_2 differs from f_1, the listener should tend to respond either that $f_2 < f_1$ or that $f_2 > f_1$; however, the listener should make errors on roughly half of these trials. Let us call errors of this sort, Type-A errors. By contrast, if a listener can discern the direction of the difference between f_2 and f_1 as soon as they can hear that $f_1 \neq f_2$, then most of the errors they make when f_2 differs from f_1 should be to respond that $f_2 = f_1$. Let us call errors of this sort, Type-B errors. As we report below, the ratio of type-A to type-B errors tends to increase with increasing PDT, suggesting, in accordance with the results of Semal and Demany (2006) and Mathias et al. (2010), that RPD-challenged listeners can often hear that $f_1 \neq f_2$ but are unable to discern the direction of the difference between them.

4.3 Methods

All methods were approved by the UCI Institutional Review Board.

4.3.1 Participants

151 undergraduate students were recruited from the Social Science Human Subjects Pool at the University of California, Irvine. All listeners had self-reported normal hearing and received course credit for participating in the study. Data were excluded from analysis if listeners scored below 5 on the Headphone Check (see Sec. 4.3.2) or if their data indicated that they did not pay attention to the tasks (i.e., they continuously pressed the same button response for half a task's trials). As a result, data for 99 listeners were analyzed for this study.

57 listeners reported having at least one year of formal musical training. Within this group of listeners, the mean number of years of musical training was 4.39 (standard deviation: 4.32).

4.3.2 Procedure

The experiment took place online at https://pitchdiffrove.web.app/ for participants to begin anytime and at their own pace. They were instructed to find a quiet room and wear headphones or earbuds for the entirety of the experiment. Listeners were free to adjust volume to their comfort level. Sampling rate of stimulus presentation was adjusted according to the sampling rate of the participant's device. If the sampling rate was outside the range of 44100 to 48000 samples/s (which would be unusual for a typical computer), then the participant was instructed to switch devices. The sampling rate was 44100 samples/s for 37 participants and 48000 samples/s for 62 participants. The specific sound card of each participant's device was unknown.

Headphone/earbud wear was screened at the start of the experiment via a 3-alternative-

forced choice task used by Woods et al. (2017). This task consists of 6 trials in which listeners judge which of three 200-Hz pure tones is quietest. Unknown to the listener, one tone in each trial is presented 180° out of phase across the stereo channels. This phase cancellation causes the task to be difficult over loudspeakers but easy over headphones. Woods et al. (2017) determined that listeners who score at least 5 correct trials can be assumed to be wearing headphones.

Following this test, listeners completed a brief survey to report their native language and number of years of musical training. They were then tested in the 3-task and 4 pitch-difference tasks. Task order was randomly generated for each listener.

4.3.3 3-task

Stimuli

Stimuli were tone-scrambles. Each tone-scramble contained 3 copies each of the following notes from the standard equal-tempered chromatic scale: G_5 (783.99 Hz), D_6 (1174.66 Hz), and G_6 (1567.98 Hz). In addition, major (minor) stimuli contained 3 copies of B_5 (987.77 Hz) ($B\flat_5$ (932.33 Hz)). Each individual tone was 65 ms in duration and was windowed by a raised cosine function with a 22.5-ms rise time. Thus, each stimulus lasted 0.78 sec.

Task

Before beginning the task, listeners heard two examples each of "Type 1 (Minor/Sad)" and "Type 2 (Major/Happy)" stimuli. Then, on each trial, listeners heard a single stimulus and were asked to classify it as "Type 1 (Minor/Sad)" and "Type 2 (Major/Happy)" by clicking buttons on the screen. Type 1 stimuli corresponded to a button depicting a "sad" face emoji on the left side of the screen; Type 2 stimuli corresponded to a button depicting a "happy" face emoji on the right side of the screen. Feedback ("Correct" or "Incorrect") was printed to the screen after each trial, and proportion correct was given at the end of the task. Listeners completed three blocks of 50 trials.

4.3.4 Pitch-difference tasks

Stimuli and task

Stimuli were pairs of pure tones. Each tone had a duration of 500-ms and was windowed by a raised cosine function with a 22.5-ms rise time.

The inter-stimulus interval and frequency of the first tone for each condition are listed in Table 4.1. The inter-stimulus interval was 1-s in the Fixed, Gap-1, and Same-Higher-Lower (SHL) conditions; the inter-stimulus interval was 500-ms in the Gap-0.5 condition.

In the Fixed condition, the first tone in each pair was fixed at 440 Hz. In the Gap-0.5, Gap-1, and SHL conditions, the frequency of the first tone in each pair was uniformly selected from the log frequency interval of 200-1600 Hz (a range of 3600 cents). In all cases, the maximum frequency difference was 1200 cents (1 octave), such that the second tone of all trials fell uniformly in the log frequency interval of 100-3200 Hz.

Table 4.1: The inter-stimulus interval (duration between the 2 tones in each stimulus, in ms) and frequency of the first tone (in Hz) for each of the 4 pitch-difference task conditions.

Condition	ISI (ms)	Frequency 1 (Hz)
Fixed	500	440
Gap-0.5	500	Roved: 200-1600
Gap-1	1000	Roved: 200-1600
Same-Higher-Lower	500	Roved: 200-1600

In each pitch-difference task, the listener heard two tones per trial and responded whether the second tone was higher or lower than the first tone. In the Same-Higher-Lower (SHL) task, the listener could also respond "same."

At the start of the SHL task, the listener heard 2 examples each of a "same" trial and 1 example each of a "higher" trial and a "lower" trial. At the start of the other tasks (Fixed, Gap-0.5, Gap-1), the listener heard 2 examples each of a "higher" trial and a "lower" trial.

Each task consisted of 2 blocks of 50 trials. Feedback ("Correct" or "Incorrect") was printed to the screen after each trial, and proportion correct was given at the end of each task.

How the frequencies of the two tones were determined on each trial.

The absolute frequency difference between the two tones in a given trial in the Fixed, Gap-0.5, and Gap-1 conditions was determined by two interleaved 3-down-1-up staircases. In a given staircase, task-difficulty was controlled by a parameter θ whose value was adjusted by the staircase. In Staircase 1, θ was set initially to 1 (corresponding to 100 cents); in Staircase 2, θ was set initially to 7 (corresponding to 700 cents). After each trial, if the previous three responses in Staircase 1 (Staircase 2) were correct, then θ was decreased to 0.9θ (0.7θ); otherwise θ was increased to 1.11θ (1.43θ). Frequency differences in the SHL task was determined only by Staircase 2.

The direction of the frequency difference (higher or lower) was assigned independently of the staircases. The second tone was higher in frequency than the first tone for half the trials in the Fixed, Gap-0.5, and Gap-1 conditions. In the SHL condition, the second tone was the same as the first tone for half the trials (regardless of the result of the staircase), and the remaining trials were evenly split between higher and lower trials.

4.4 Results

4.4.1 3-task

Performance in the 3-task, measured as d', was computed from the last block of 50 trials. The first 2 blocks of trials was treated as practice. If a listener was tested on n major (minor) stimuli and responded correctly on all of them, then the probability of a correct response was adjusted to n - $0.5/n$ (as suggested by Macmillan and Kaplan (1985)).

The distribution of all listeners' d' values are plotted as a histogram in Fig. 4.2 in gray bars. Similar to previous tone-scramble studies, we observe that the majority of listeners achieved d' values near 0, corresponding to chance performance, while the remaining listeners achieved high d' values near 4.1, corresponding to near-perfect performance. However, the

proportion of low-performing listeners in our sample (about 88% of participants) is much greater than what we typically observe in previous studies (about 70% of participants).

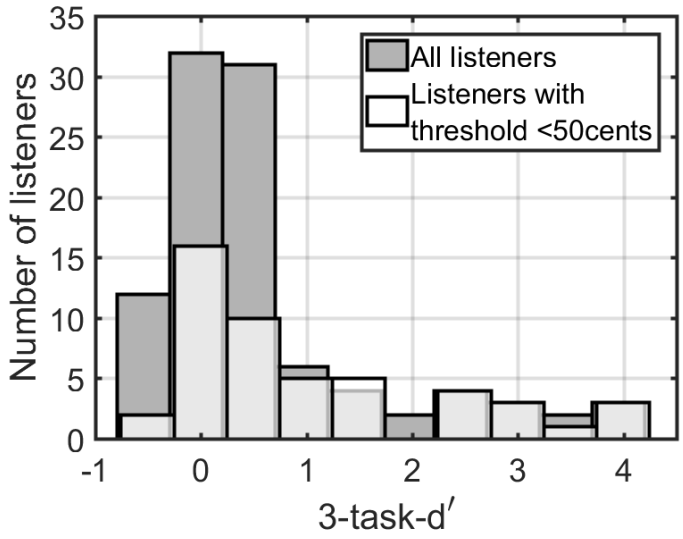

Figure 4.2: Histogram of d' values achieved on the 3-task by all listeners (gray bars). The white bars (slightly shifted to the right for visualization purposes) represent the distribution of d' values when listeners who achieved Gap-0.5 thresholds above 50 cents are excluded.

4.4.2 Pitch-difference tasks

Pitch difference threshold (PDT), which we defined as the absolute value of the pitch-difference (in cents) at which a listener achieves 80% accuracy, was estimated for the last 70 trials of each listener for each of the Fixed, Gap-0.5, and Gap-1 tasks. We fit the following Weibull function to the data from each listener in each task:

$$\Psi(D) = 0.5 + 0.48\left[1 - \exp\left(-\left(\frac{D}{A}\right)^{B}\right)\right] \tag{4.1}$$

where D is the absolute value of the difference between f_2 and f_1 on a given trial, and A and B are the Weibull function threshold and steepness parameter respectively. The reader will note that

1. $\Psi(0) = 0.5$, reflecting the fact that chance performance is 0.5 in this task, and

2. $\Psi(D) \rightarrow 0.98$ as D grows large. This limit on probability correct is intended to cover the possibility of "finger errors," i.e., incorrect responses that occur even when the listener knows the correct answer.

We use the maximum likelihood estimate of A as the listener's threshold. For $D = A$, proportion correct is 0.8034; thus the PDT's reported here are predicted to support performance around 80% correct.

To check the accuracy of the PDT estimate for each listener, we used Markov chain Monte Carlo simulation to derive samples from the posterior joint density characterizing the parameters A and B. If the PDT of a listener was lower than 50 cents (a quarter-tone), the 100 trials of data obtained from that listener typically sufficed to tightly constrain the estimate of A (i.e., the Bayesian credible interval around A was small). However, if the PDT of a listener was higher than 50 cents, this was often not true. The data from these low-performing listeners was often very ragged, and the values visited by their staircases tended to range widely; consequently, in such cases, the credible interval around A sometimes spanned several orders of magnitude. Nonetheless, the maximum-likelihood Weibull function estimates generally did a reasonable job of capturing the overall trends even in the most aberrant data sets. Thus, although it would be a mistake to take the PDT estimate for a given, low-performing listener too seriously, in each case, the estimated PDT appears sensible based on the available data.

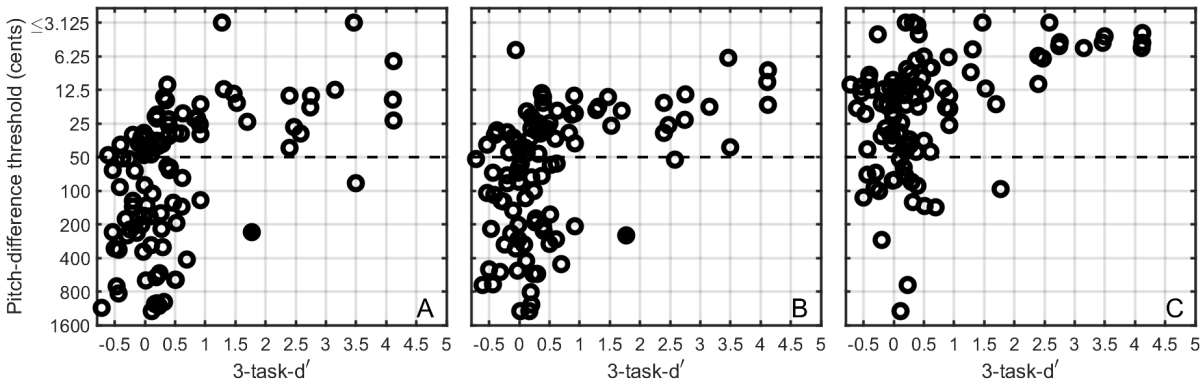

Figure 4.3: Scatterplot of pitch-difference-threshold from the (A) Gap-0.5 task, (B) Gap-1 task, and (C) Fixed task as a function of 3-task-d'. Pitch-difference-thresholds are plotted (on a log scale) with values decreasing from bottom to top to reflect increasingly good performance. The dashed line is at 50 cents (half-a-semitone). The outlier in (A) and (B) is plotted as a filled circle.

Fig. 4.3 plots each listener's d' on the 3-task against their threshold for the Fixed, Gap-0.5, and Gap-1 pitch-difference tasks. Thresholds are plotted along the y-axis of each plot on a log scale decreasing from bottom to top to reflect increasing levels of performance. The horizontal dotted line at 50 cents marks sensitivity to half a semitone; listeners who are plotted below this line struggle to judge the direction (higher or lower) of 2 tones unless the tones are more than half a semitone apart. Listeners whose threshold exceeds 100 cents and whose d' is greater than 1 are plotted in filled circles - these listeners are considered to be outliers. All listeners with thresholds greater than 50 cents (with the exception of the outliers) achieve $d' < 1$. The majority of the listeners with thresholds below 50 cents are centered around d'=0 with few listeners achieving high d', unlike in the experiment by Mann (2014) which revealed a mostly uniform distribution of d' for listeners with thresholds below 50 cents. This inconsistency can most likely be attributed to the greater proportion of listeners observed to perform near chance on the 3-task in the current experiment.

The lighter bars in Fig. 4.2 reveal the distribution of 3-task d' values when the listeners who achieved Gap-0.5 thresholds above 50 cents are excluded. Similar to the results from Mann (2014), the large peak around d'=0 is greatly reduced, showing that most listeners

whose thresholds exceed 50 cents are performing around chance on the 3-task. The reduction of this peak is not as dramatic as in the results of Mann (2014), probably because a greater proportion of listeners in Experiment 2 performed near a d' of 0.

The reader will observe in Fig. 4.3 that the relationship between d' and threshold is similar for the Gap-0.5 and Gap-1 tasks: Listeners whose d' values fall below 1 achieve a wide range of thresholds both below and above the 50-cent line; as d' increases, thresholds are mostly less than 50 cents. This pattern also occurs in the Fixed condition; however, threshold values are shifted such that a greater proportion of the listeners whose d' values fall below 1 achieve thresholds less than 50 cents. The histograms plotted in Fig. 4.4 further illustrate the relationship between the thresholds achieved in the Fixed condition and the two roved conditions. The log ratio of thresholds for Fixed vs. Gap-0.5 (gray bars, mean = -0.58, standard error = 0.05) and Fixed vs. Gap-1 (white bars, mean = -0.57, standard error = 0.04) both appear mostly normal with a nearly identical mean ratio. Therefore, most listeners achieve Fixed thresholds that are nearly half of their threshold in either roved condition. This effect was slightly greater in the Fixed-to-Gap-0.5 comparison for listeners who achieved Gap-0.5 threshold below 50 cents. The mean log ratio of Fixed to Gap-0.5 threshold for these low-threshold listeners was -0.48, while the mean log ratio for high-threshold listeners (Gap-0.5 threshold over 50 cents) was -0.69. A two-tailed, two-sample t-test of the null hypothesis that the log ratio of Fixed to Gap-0.5 thresholds of low-threshold listeners and the log ratio of Fixed to Gap-0.5 thresholds of high-threshold listeners come from distributions with equal means, assuming unequal variances, yielded $t(93.2) = 2.34$, $p = 0.02$. However, when applied to the Fixed to Gap-1 threshold comparison, the results of the t-test was insignificant, yielding $t(96.72) = 0.21$, $p = 0.84$.

Performance on the Same-Higher-Lower task was analyzed separately from the other pitch-difference tasks. Unlike the other tasks, the two tones presented on a given trial in the SHL task could possibly be the same in pitch, and listeners were required to judge whether the tones were the same, higher, or lower. Therefore, this task tested listeners'

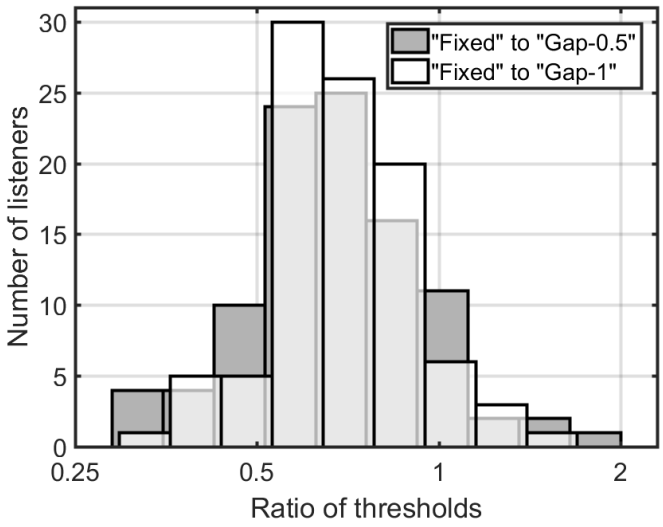

Figure 4.4: Histogram of ratios (plotted on a log scale) of pitch-difference thresholds for Fixed vs. Gap-0.5 (gray bars, mean = -0.58, standard error = 0.05) and Fixed vs. Gap-1 (white bars, mean = -0.57, standard error = 0.04). The distribution for Fixed vs. Gap-1 (white bars) is slightly shifted to the right for visualization purposes. Both distributions appear mostly normal with a nearly identical mean ratio.

ability to judge not only whether two tones have a pitch difference, but also whether that pitch difference occurs in a higher or lower direction.

Each listener's direction confusability statistic, or the ratio of Type-A errors (related to pitch-*direction*) to Type-B errors (related to pitch-*difference*), was computed using the trials following their 3rd error (to allow their staircase to stabilize):

$$\text{direction confusability} = \frac{A + B}{C + 0.5} \tag{4.2}$$

where A is the number of "Lower" trials on which a listener incorrectly responds "Higher," B is the number of "Higher" trials on which the listener incorrectly responds "Lower," and C is the number of "Lower" and "Higher" trials on which the listener incorrectly responds "Same." A listener who is sensitive to pitch difference but has difficulty identifying the direction should yield a higher direction confusability statistic, because they correctly recognize

when two pitches are different but make errors when judging whether a pitch is higher or lower.

Fig. 4.5 plots each listener's direction confusability statistic against their pitch-difference-threshold on the Gap-0.5 task. We adjusted the direction confusability value of one outlier (plotted as a filled circle) from 8.67 to 2. All listeners who achieve a statistic of 0 have thresholds less than 75 cents. The majority of listeners whose threshold is less than 100 cents achieve statistics less than 0.5. The distribution spreads out for listeners whose thresholds exceed 100 cents, uniformly ranging from 0 to 2.

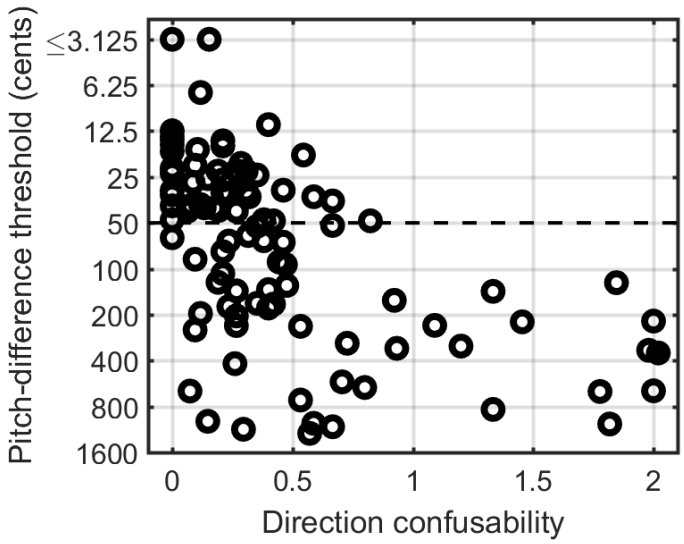

Figure 4.5: Relationship between direction confusability and pitch-difference-threshold. Pitch-difference-thresholds are plotted (on a log scale) with values decreasing from bottom to top to reflect increasingly good performance. The dashed line is at 50 cents (half-a-semitone). The direction confusability value of one outlier (plotted as a filled circle) was adjusted from 8.67 to 2.

4.4.3 Effects of musical training

Fig. 4.6A plots each listener's self-reported years of musical training against their 3-task-d' ($r = 0.41, p < 0.01$). Fig. 4.6B plots years of musical training against each listener's pitch-difference-threshold (in cents) on the Gap-0.5 task ($r = -0.34, p < 0.01$). The correlation between years of musical training and threshold on the Fixed task (not pictured) was lower

$(r = -0.21, p < 0.05)$.

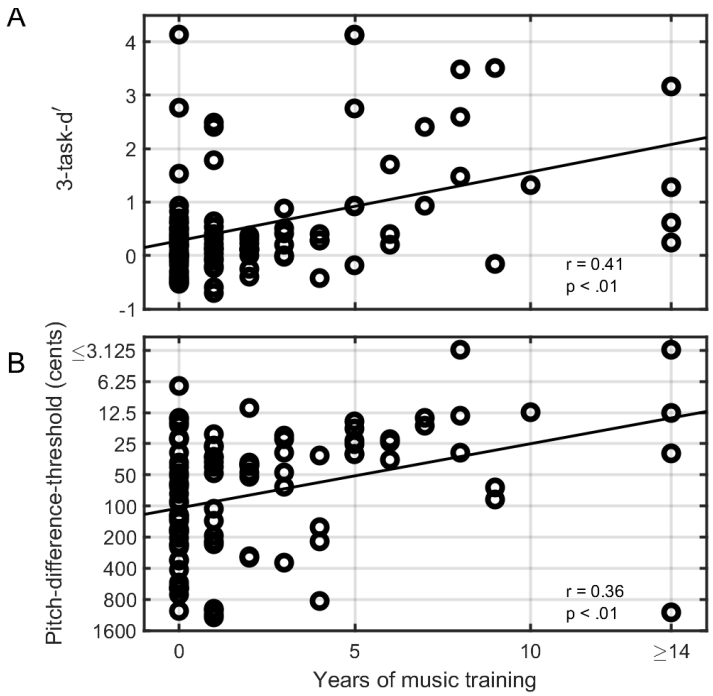

Figure 4.6: Relationship between years of musical training and (a) 3-task-d' and (b) pitch-difference threshold.

In the group of 21 listeners with at least 5 years of musical training, 9 listeners had d' below 1. Two of these 9 listeners, in addition to a listener with a d' of 3.5 and 9 years of musical training, had threshold higher than 50 cents. Among the four listeners with at least 14 years of training, only 1 achieved a d' above 3, and 3 achieved thresholds less than 50 cents.

Among the 59 listeners with fewer than 2 years of musical training, a listener with no years of training achieved the highest d' (4.12) and the lowest threshold (6.98). Five of the 6 listeners who achieved d' above 3 had at least 5 years of musical training.

4.5 Discussion

Consistent with previous studies (Chubb et al., 2013; Dean & Chubb, 2017; Mednicoff et al., 2018; Ho & Chubb, 2020), most listeners achieved low d' values around chance performance

on the 3-task. However, we did not observe the usual 70%-30% distribution of low-vs-high performance from previous experiments. A reasonable explanation is that the online format of the current study resulted in increased inattention and non-compliance among listeners. We had attempted to minimize non-compliance by conducting the headphone check test at the start of the experiment to filter out listeners who did not follow instructions about wearing headphones, and the number of these listeners was quite high (46 out of 151). Furthermore, the distinct pattern of thresholds between the roved (Gap-0.5 and Gap-1) and Fixed conditions of the pitch-difference tasks suggests that the remaining listeners were paying attention during these tasks. A more likely explanation of the greater proportion of low performers on the 3-task can be attributed to our usage of fewer notes in our 3-task stimuli (12 notes compared to 32 notes in previous iterations). Preliminary data collected in parallel with this study suggests that the 12-note stimuli are more difficult to discriminate than the 32-note stimuli; thus, we speculate that several low-performing listeners in the current study would attain more correct trials in the 32-note version.

Similar to the results from Mann (2014), listeners (with the exception of some outliers) cannot achieve high performance on the 3-task if their PDT exceeds 50 cents. We also observe that having a low PDT does not guarantee high performance on the 3-task.

As expected, PDTs were elevated when the f_1 in each trial of a task was roved (as in the Gap-0.5 and Gap-1 conditions), resembling findings from previous studies (Amitay, Hawkey, & Moore, 2005; Demany & Semal, 2005; Harris, 1952; Jesteadt & Bilger, 1974; Mathias et al., 2010). Mathias et al. (2010) found that novice listeners (without prior experience with pitch-difference tasks) tended to demonstrate impaired pitch-direction identification for wide frequency-roving ranges (3102 cents) compared to medium (310 cents) or narrow (31 cents) roving ranges. The roving range in the current experiment was 3600 cents. Mathias et al. (2010) hypothesize that frequency-roving increases stimulus uncertainty, contributing to internal noise and therefore impairing listeners' performance. This effect possibly has a greater influence on the listeners who have higher PDTs. Connecting this idea to 3-task

performance, we speculate that listeners with low 3-task-d' struggle on the 3-task as a result of having not only higher PDT but also greater susceptibility to informational masking. In other words, these listeners may experience greater difficulty to perceive the target tones of the 3-task stimuli because they are less able to suppress the internal noise exerted by the non-target tones.

In tasks with a fixed f_1 (i.e., the Fixed condition) which measure sensitivity to deviations in frequency from a fixed standard, the listener can combine information across trials to construct (and refresh) a durable "remembered standard" against which to compare new frequencies. Roved-frequency tasks, in contrast, require the listener to retain a distinct memory of the current f_1 on any given trial. The manipulation of inter-stimulus interval in the current experiment did not appear to drastically affect listeners' perception of pitch differences. However, the results from the comparisons of Fixed thresholds to either roved conditions' thresholds imply that ISI had some (but very subtle) impact on pitch-comparison threshold. In particular, low-PDT listeners (Gap-0.5 PDT $<$ 50 cents) and high-PDT listeners (Gap-0.5 PDT $>$ 50 cents) differed in their log ratio of Fixed PDT to Gap-0.5 PDT, but this was not the case for the log ratio of Fixed PDT to Gap-1 PDT.

The results from the SHL task reveal that listeners seem to require a larger pitch difference to recognize both pitch difference and pitch direction. Listeners achieved a direction confusability statistic of 0 only if they did not make pitch-direction errors (responding "Higher" to a "Lower" trial or responding "Lower" to a "Higher" trial). All listeners who fell into this group had a threshold less than 75 cents. Some listeners with a higher threshold were able to achieve low statistics near 0. We speculate that these listeners are somewhat sensitive to pitch direction but require a larger pitch difference to identify the direction. The listeners whose statistic exceeds 1 all have thresholds greater than 100 cents.

The results for musical training reiterate the findings from previous studies of the 3-task that musicianship does not directly predict sensitivity to major-vs-minor. Many listeners with little or no musical training achieved high 3-task-d' and PDT below 50 cents. Thus,

skill on the 3-task and PDT are possibly preexisting abilities that manifest independently of musical training.

In summary, this experiment demonstrates that PDT may determine a listener's ability to do well on the 3-task. Specifically, a PDT below 50 may be necessary to achieve high performance on the 3-task. Not all listeners with low PDT do well on the 3-task, but perhaps these listeners are capable of improving if they receive additional intervention. The type of intervention in question, whether it be general music training or intensive practice on the 3-task, is unclear at this time.

Chapter 5

General Discussion and Future Directions

This dissertation investigated the source of scale-sensitivity through various perspectives. Chapter 2 revealed that changing the temporal structure of stimuli in the 3-task can make the task easier for listeners. This result suggests that listeners with high scale-sensitivity (who already perform near perfect on the standard 3-task) may possess more effective grouping abilities to hear target note differences. Chapter 3 showed that the same processing resources may underlie scale-sensitivity and speech prosody perception, and these resources are largely unaffected by musical training. Chapter 4 suggested that one of these processing resources might be pitch-difference threshold: only the listeners with pitch-difference thresholds below 50 cents can achieve high scale-sensitivity.

Considering the established link between major/minor with happy/sad emotions, it is surprising that most listeners struggle to differentiate major-vs-minor, performing around chance on the 3-task. If we were to extend this phenomenon to the greater population, however, it would seem unlikely for most of the population to be unable to differentiate between happy and sad music, and potentially be unable to separate happy from sad speech. Natural music typically contains additional cues (such as tempo, volume, and lyrical content)

that convey the intended mood, while tone-scramble stimuli are stripped down to vary only by pitch. It makes sense, then, that pitch-difference threshold might be the factor that determines a listener's sensitivity to the different tone-scramble stimuli. A future study could focus on testing the listeners with low pitch-difference threshold (below 50 cents) and low 3-task-d'. If a low pitch-difference threshold is the main precondition for high 3-task performance, then this group of listeners can potentially benefit from additional practice on the 3-task or some other form of targeted training.

Altogether, these studies present evidence that scale-sensitivity is mostly an inherent skill, rather than a product of musical training. Although taking music lessons can potentially improve one's musical (and non-musical) abilities in various ways, recent studies emphasize the importance of taking preexisting musicality into account when attributing a listener's performance on any task to their musical background (e.g., Correia et al. (2020); Kragness et al. (2020)). In future studies of the 3-task, we might administer questionnaires that objectively assess musical ability, such as the Ollen Musical Sophistication Index (OMSI) (Ollen, 2006), Goldsmiths Musical Sophistication Index (Gold-MSI) (Müllensiefen, Gingras, Musil, & Stewart, 2014), or Profile of Music Perception Skills (PROMS) (Law & Zentner, 2012), to test whether scale-sensitivity correlates with scores on these surveys. We also hope that other researchers in music cognition will recognize the value in measuring listeners' scale-sensitivity in their own studies, since scale-sensitivity (rather than musical training) may possibly underlie performance on their target task.

Future studies can also explore whether scale-sensitivity actually translates to emotion perception by comparing scale-sensitivity to performance on tasks that require the listener to judge the perceived emotions of different types of music or speech. Studies can also investigate whether the shared connection between scale-sensitivity and speech perception extends to other types of speech processing that largely involve tracking pitch over time, such as the ability to track a speaker's voice in multi-talker speech, or the ability to identify tones in a tonal language. Finally, studies can also analyze scale-sensitivity from a neural

perspective, such as through EEG, to explore whether scale-sensitivity is represented through different electrophysiological responses (thus reflecting that listeners use different strategies to classify major-vs-minor).

Bibliography

Adler, S. A., Comishen, K. J., Wong-Kee-You, A. M. B., & Chubb, C. (2020, https://doi.org/10.1121/10.0001349 DOI: 10.1121/10.0001349). Sensitivity to major versus minor musical modes is bimodally distributed in young infants. *Journal of the Acoustical Society of America*, *147*, 3758-3764.

Amitay, S., Hawkey, D. J., & Moore, D. R. (2005). Auditory frequency discrimination learning is affected by stimulus variability. *Perception & psychophysics*, *67*(4), 691–698.

Asaridou, S. S., & McQueen, J. M. (2013). Speech and music shape the listening brain: evidence for shared domain-general mechanisms. *Frontiers in psychology*, *4*, 321.

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

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

Bidelman, G. M., Hutka, S., & Moreno, S. (2013). Tone language speakers and musicians share enhanced perceptual and cognitive abilities for musical pitch: evidence for bidirectionality between the domains of language and music. *PloS one*, *8*(4).

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.

Boersma, P., & Weenink, D. (2017). Praat: Doing phonetics by computer [computer program]. , Version 6.0.34, retrieved 11 October 2017 from http://www.praat.org/.

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

Bowling, D. L., Sundararajan, J., Han, S., & Purves, D. (2012). Expression of emotion in eastern and western music mirrors vocalization. *PLoS One*, *7*(3), e31942.

Burnham, B. R., Long, E., & Zeide, J. (2020). Pitch direction on the perception of major and minor modes. *Attention, Perception, & Psychophysics*, 1–16.

Buss, E., Taylor, C. N., & 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., ... Kowalski, E. (2013). Bimodal distribution of performance in discriminating major/minor modes.

Journal of the Acoustical Society of America, *134*(4), 3067-3078.
Correia, A. I., Castro, S. L., MacGregor, C., Mullensiefen, D., Schellenberg, E. G., & Lima, C. F. (2020). Enhanced recognition of vocal emotions in individuals with naturally good musical abilities. *Emotion*, *http://dx.doi.org/10.1037/emo0000770*, 1-13.
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., & Sterling, R. S. (1988). Developmental change in the understanding of affective meaning in music. *Motivation and Emotion*, *12*, 399-413.
Curtis, M. E., & Bharucha, J. J. (2010). The minor third communicates sadness in speech, mirroring its use in music. *Emotion*, *10*(3), 335-348.
Dean, T., & Chubb, C. (2017). Scale-sensitivity: A cognitive resource basic to music perception. *Journal of the Acoustical Society of America*, *142*(3), 1432-1440.
Degé, F., & Schwarzer, G. (2011). The effect of a music program on phonological awareness in preschoolers. *Frontiers in psychology*, *2*, 124.
Demany, L., & Semal, C. (2005). The slow formation of a pitch percept beyond the ending time of a short tone burst. *Perception & psychophysics*, *67*(8), 1376–1383.
Deutsch, D. (1982). Organizational processes in music. In *Music, mind, and brain* (pp. 119–136). Springer.
Deutsch, D., Henthorn, T., Marvin, E., & Xu, H. (2006). Absolute pitch among american and chinese conservatory students: Prevalence differences, and evidence for a speech-related critical period. *Journal of the Acoustical Society of America*, *119*, 719-722.
Dowling, W. J. (1973). Rhythmic groups and subjective chunks in memory for melodies. *Perception & Psychophysics*, *14*(1), 37–40.
Farmer, E., Jicol, C., & Petrini, K. (2020). Musicianship enhances perception but not feeling of emotion from others' social interaction through speech prosody. *Music Perception*, *37*(4), 323–338.
Fujioka, T., Trainor, L. J., Ross, B., Kakigi, R., & 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., & Pantev, C. (2005). Automatic encoding of polyphonic melodies in musicians and nonmusicians. *Journal of cognitive neuroscience*, *17*(10), 1578-1592.
Gagnon, L., & Peretz, I. (2003). Mode and tempo relative contributions to "happy-sad" judgements in equitone melodies. *Cognition and Emotion*, *17*(1), 25-40.
Gandour, J., Wong, D., Hsieh, L., Weinzapfel, B., Lancker, D. V., & Hutchins, G. D. (2000). A crosslinguistic pet study of tone perception. *Journal of cognitive neuroscience*, *12*(1), 207–222.
Gerardi, G. M., & Gerken, L. (1995). The development of affective responses to modality and melodic contour. *Music Perception*, *12*(3), 279-290.
Glushko, A., Steinhauer, K., DePriest, J., & Koelsch, S. (2016). Neurophysiological correlates of musical and prosodic phrasing: Shared processing mechanisms and effects of musical

expertise. *PLoS ONE*, *11*(5).

Gordon, R. L., Magne, C. L., & Large, E. W. (2011). Eeg correlates of song prosody: a new look at the relationship between linguistic and musical rhythm. *Frontiers in Psychology*, *2*(352), doi: 10.3389/fpsyg.2011.00352.

Gupta, P., Lipinski, J., Abbs, B., Lin, P.-H., Aktunc, E., Ludden, D., ... Newman, R. (2004). Space aliens and nonwords: Stimuli for investigating the learning of novel word-meaning pairs. *Behavior Research Methods, Instruments, & Computers*, *36*(4), 599–603.

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

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

Hambrick, D. Z., & Tucker-Drob, E. M. (2015). The genetics of music accomplishment: Evidence for gene–environment correlation and interaction. *Psychonomic bulletin & review*, *22*(1), 112–120.

Harris, J. D. (1952). The decline of pitch discrimination with time. *Journal of experimental psychology*, *43*(2), 96.

Heffner, C. C., & Slevc, L. R. (2015). Prosodic structure as a parallel to musical structure. *Frontiers in psychology*, *6*, 1962.

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., & Chubb, C. (2020, https://doi.org/10.1121/10.0001398). How rests and cyclic sequences influence performance in tone-scramble tasks. *Journal of the Acoustical Society of America*, *147*, 3859-3870.

Hoch, L., Poulin-Charronnat, B., & Tillman, B. (2011). The influence of task-irrelevant music on language processing: syntactic and semantic structures. *Frontiers in Psychology*, *2*(112), doi:10.3389/fpsyg.2011.00112.

Hoel, P. G., Port, S. C., & Stone, C. J. (1971). *Introduction to statistical theory.* Boston, MA: Houghton-Mifflin.

Huron, D. (2008). A comparison of average pitch height and interval size in major-and minor-key themes: Evidence consistent with affect-related pitch prosody. *Empirical Musicology Review*, *3*(2), 59-63.

Huron, D., & Davis, M. J. (2012). The harmonic minor scale provides an optimum way of reducing average melodic interval size, consistent with sad affect cues. *Empirical Musicology Review*, *7*(3-4), 103-117.

Jesteadt, W., & Bilger, R. C. (1974). Intensity and frequency discrimination in one-and two-interval paradigms. *The Journal of the Acoustical Society of America*, *55*(6), 1266–1276.

Johnsrude, I. S., Penhune, V. B., & Zatorre, R. J. (2000). Functional specificity in the right human auditory cortex for perceiving pitch direction. *Brain*, *123*(1), 155–163.

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

770-814.
Kastner, M. P., & Crowder, R. G. (1990). Perception of the major/minor distinction: Iv. emotional connotations in young children. *Music Perception*, *8*(2), 189-202.
Knopoff, L., & Hutchinson, W. (1983). Entropy as a measure of style: The influence of sample length. *Journal of Music Theory*, *27*(1), 75–97.
Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T., & Friederici, A. D. (2004). Music, language and meaning: Brain signatures of semantic processing. *Nature Neuroscience*, *7*(3), 302-307.
Kragness, H. E., Swaminathan, S., Cirelli, L. K., & Schellenberg, E. G. (2020). Individual differences in musical ability are stable over time in childhood. *Developmental Science*.
Kraus, N., Slater, J., Thompson, E. C., Hornickerl, J., & Strait, D. L. (2014). Music enrichment programs improve the neural encoding of speech in at-risk children. *The Journal of Neuroscience*, *34*(36), 11913-11918.
Krishnan, A., Xu, Y., Gandour, J., & Cariani, P. (2005). Encoding of pitch in the human brainstem is sensitive to language experience. *Cognitive Brain Research*, *25*(1), 161–168.
Krumhansl, C. (1990). *Oxford psychology series.* Cognitive foundations of musical pitch. New York, NY, US: Oxford University ....
Krumhansl, C. L., & Kessler, E. J. (1982). Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychological Review*, *89*(4), 334-368.
Lantz, M. E., & Cuddy, L. L. (1998). Total and relative duration as cues to surface structure in music. *Canadian Acoustics*, *26*(3), 56–57.
Law, L. N., & Zentner, M. (2012). Assessing musical abilities objectively: Construction and validation of the profile of music perception skills. *PloS one*, *7*(12), e52508.
Leaver, A. M., & Halpern, A. R. (2004). Effects of training and melodic features on mode perception. *Music Perception*, *22*, 117-143.
Lima, C. F., & Castro, S. L. (2011). Speaking to the trained ear: musical expertise enhances the recognition of emotions in speech prosody. *Emotion*, *11*(5), 1021.
Loui, P., Kroog, K., Zuk, J., Winner, E., & Schlaug, G. (2011). Relating pitch awareness to phonemic awareness in children: implications for tone-deafness and dyslexia. *Frontiers in Psychology*, *2*(11), doi: 1.10.3389/fpsyg.2011.00111.
Macmillan, N. A., & 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.
Mann, D. S. (2014). *Processing stimuli over time: Musical modes and audiovisual binding* (Unpublished doctoral dissertation). University of California, Irvine.
Margulis, E. H., & Simchy-Gross, R. (2016). Repetition enhances the musicality of randomly generated tone sequences. *Music Perception: An Interdisciplinary Journal*, *33*(4), 509–514.
Marie, C., Delogu, F., Lampis, G., Belardinelli, M. O., & 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., & Besson, M. (2010). Musicians and the metric structure of words. *Journal of cognitive neuroscience*, *23*(2), 294-305.

Mathias, S. R., Micheyl, C., & Bailey, P. J. (2010). Stimulus uncertainty and insensitivity to pitch-change direction. *The Journal of the Acoustical Society of America*, *127*(5), 3026–3037.

Mednicoff, S., Mejia, S., Rashid, J., & 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.

Meyer, L. B. (2008). *Emotion and meaning in music*. University of chicago Press.

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

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review*, *63*(2), 81.

Morrill, T. H., Devin, J. D., Dilley, L. C., & 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.

Mosing, M. A., Madison, G., Pedersen, N. L., Kuja-Halkola, R., & Ullén, F. (2014). Practice does not make perfect: no causal effect of music practice on music ability. *Psychological science*, *25*(9), 1795–1803.

Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The musicality of non-musicians: an index for assessing musical sophistication in the general population. *PloS one*, *9*(2), e89642.

Ollen, J. E. (2006). *A criterion-related validity test of selected indicators of musical sophistication using expert ratings* (Unpublished doctoral dissertation). The Ohio State University.

Ott, C. G., Langer, N., Oechslin, M., & Jancke, L. (2011). Processing of voiced and unvoiced acoustic stimuli in musicians. *Frontiers in Psychology*, *2*(195), doi: 10.3389/fpsyg.2011.00195.

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

Parbery-Clark, A., Skoe, E., Lam, C., & 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., & 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., Iversen, J. R., & 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.

Peretz, I. (2002). Brain specialization for music. *Neuroscientist*, *8*(4), 372.

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

Plomp, R., & Levelt, W. J. M. (1965). Tonal consonance and critical bandwidth. *The journal of the Acoustical Society of America*, *38*(4), 548–560.

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

Roncaglia-Denissen, M. P., Bouwer, F. L., & Honing, H. (2018). Decision making strategy and the simultaneous processing of syntactic dependencies in language and music. *Frontiers in Psychology*, *9*, DOI: 10.3389/fpsyg.2018.00038.

Rosenthal, M. A., & Hannon, E. E. (2016). Cues to perceiving tonal stability in music: The role of temporal structure. *Music Perception: An Interdisciplinary Journal*, *33*(5), 601–612.

Sallat, S., & Jentschke, S. (2015). Music perception influences language acquisition: Melodic and rhythmic-melodic perception in children with specific language impairment. *Behavioral Neurology*.

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

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

Semal, C., & Demany, L. (2006). Individual differences in the sensitivity to pitch direction. *Journal of the Acoustical Society of America*, *120*(6), 3907-3915.

Smith, N. A., & Schmuckler, M. A. (2004). The perception of tonal structure through the differentiation and organization of pitches. *Journal of experimental psychology: human perception and performance*, *30*(2), 268.

Strait, D. L., & Kraus, N. (2011). Can you hear me now? musical training shapes functional brain networks for selective auditory attention and hearing speech in noise. *Frontiers in Psychology*, *2*(113), doi: 10.3389/fpsyg.2011.00113.

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

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

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

Trainor, L. J., Tsang, C. D., & Cheung, V. H. (2002). Preference for sensory consonance in 2-and 4-month-old infants. *Music Perception: An Interdisciplinary Journal*, *20*(2), 187–194.

Tramo, M. J., Shah, G. D., & Braida, L. D. (2002). Functional role of auditory cortex in frequency processing and pitch perception. *Journal of neurophysiology*, *87*(1), 122–139.

Trimmer, C. G., & Cuddy, L. L. (2008). Emotional intelligence, not music training, predicts recognition of emotional speech prosody. *Emotion*, *8*(6), 838.

Tymoczko, D. (2011). *A geometry of music–harmony and counterpoint in the extended common practice*. Oxford University Press.

Wilks, S. S. (1944). *Mathematical statistics*. Princeton University Press.

Wong, P. C., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature neuroscience*,

10(4), 420–422.

Youngblood, J. E. (1958). Style as information. *Journal of Music Theory*, *2*(1), 24–35.

Zuk, J., Benjamin, C., Kenyon, A., & 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.