Title
Scrambled Tones From Another Dimension: Probing the Mechanisms Underlying Tonality Perception.

Author
Dean, Tyler

Publication Date
2015

UNIVERSITY OF CALIFORNIA,
IRVINE

Scrambled Tones From Another Dimension: Probing the Mechanisms Underlying Tonality Perception.

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Psychology

by

Tyler Dean

Dissertation Committee:
Professor Charles Chubb, Chair
Professor Virginia Richards
Professor Kourosh Saberi
Associate Professor Charles E. Wright

2015

DEDICATION

To

my mother and father, Jan and Larry

You have always been so supportive. I am so grateful to have had you both in my life. I would not be where I am today if not for you both. In particularly, your love for music and your insistence that it be a large part of my life has not only made my life so much more fulfilling but also led me to pursue this path in my studies.

TABLE OF CONTENTS

LIST OF FIGURES

Page

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my committee chair and advisor, Charles Chubb. Your unwavering positivity is constantly inspiring, and I feel so lucky to have studied with you. Your help has been not just crucial for this document, but for my education as a person. Since graduate school began, my way of thinking about the world has changed, not just for academics, but for so much of what I do day to day. I am so grateful for the opportunity to have learned what I have from you, and I hope one day I can be a fraction of the teacher for someone else that you were to me.

I would like to thank my committee members, Professor Charles E. Wright, Professor Kourosh Saberi, and Professor Virginia Richards. All of your guidance and effort has been so helpful in this entire process.

In addition, a thank you to Debbie Emerson of Amador Valley High School. Without your incredible enthusiasm and wonderful teaching, I may not have ever fallen in love with Psychology.

Financial support was provided by the University of California, Irvine School of Social Sciences Graduate teaching assistant award and Dean's fellowship. Thank you so much to the department and school for giving me the chance to study here and learn so very much.

CURRICULUM VITAE

Tyler Dean
Department of Cognitive Sciences
University of California, Irvine
Irvine, CA 92697
Email: deant@uci.edu

EDUCATION

2009 B.A. in Psychology, University of California, Irvine

2014 M.A. in Psychology, University of California, Irvine

2015 Ph.D. in Psychology
University of California, Irvine

PUBLICATIONS

Chubb, C., Dickson, C. A., Dean, T., Fagan, C., Mann, D. S., Wright, C. E., ... & Kowalsky, E. (2013). Bimodal distribution of performance in discriminating major/minor modes. *The Journal of the Acoustical Society of America*, *134*(4), 3067-3078.

ACADEMIC AWARDS, HONORS, AND FELLOWSHIPS

2014 Associate Dean's Fellowship from the University of California, Irvine School of Social Sciences.

TEACHING EXPERIENCE

2010-2015 Teaching Assistant at the University of California, Irvine for fourteen quarters

2013 Guest Lecture: Higher Order Perception, Psychology Fundamentals, University of California, Irvine. Irvine, California.

2014 Teacher, Introduction to Psychology, University of California, Irvine

TEACHING ASSISTANTSHIPS

Spring 2015 Psychology of Sleep and Consciousness – Nicole Gage

Winter 2015 Industrial/Organizational Psychology – Jacklyn Lewis

Spring 2014	Introduction to Psychology – John Hagedorn
Winter 2014	Psychology Fundamentals – Mark Styvers
Fall 2013	Vision – George Sperling
Spring 2013	Psychology Fundamentals – Bruce Berg
Winter 2012	Writing About Memory – Charles Chubb
Fall 2012	Critical Reasoning – Jeremy Heis
Spring 2012	Probability and Statistics in the Social Sciences – Nicholas Noviello
Winter 2012	Probability and Statistics in the Social Sciences – Nicholas Noviello
Fall 2011	Probability and Statistics in the Social Sciences – Nicholas Noviello
Spring 2011	Teaching Assistant for General Statistics and Probability – Paul Shirey
Winter 2011	Introduction to Psychology – John Hagedorn
Fall 2010	Introduction to Psychology – Christine Lofgren

ABSTRACT OF THE DISSERTATION

Scrambled Tones From Another Dimension: Probing the Mechanisms Underlying Tonality Perception.

By

Tyler Dean

Doctor of Philosophy in Psychology

University of California, Irvine, 2015

Professor Charles Chubb, Chair

Music is capable of eliciting a wide variety of sensations and emotions. Current models of tonality perception assume that musical elements (tones, chords, phrases) activate multiple mechanisms in the brain that give rise to these sensations and emotions. Previous research has primarily used subjective judgments from trained musicians to derive the fundamental dimensions of tonality. Importantly, this research has used feedback-free experimental paradigms in which the listener must be trusted as the expert witness to ground truth. The current research, by contrast, uses feedback-driven tasks to measure the sensitivity of listeners to variations in the physical properties of musical stimuli. In this dissertation, I investigate several notions of dimensionality in tonality specifically looking at sensitivity in both experts and novices. We focus on the sensitivity of listeners to variations in the global statistics of brief, rapid, randomly ordered sequences of musical tones called tone-scrambles. Previous research (Chubb et al., 2013) has shown that (1) most listeners (roughly 70%) cannot hear any difference between tone-scrambles that differ only in degree three of the major vs. minor diatonic scale, and (2) the other listeners

are highly sensitive to this difference. In chapter 1, I show that degree six of the diatonic scale splits the population in the same way as scale degree three. In chapter two, a large sample of listeners is tested in each of the “3” and “6” tasks previously investigated along with three other similar tasks, in each of which the tones to be discriminated differ by a semitone. Performance in all five tasks is strongly correlated showing that a single mechanism predominates in enabling judgments of semitone differences. In chapter three, I take a much closer look at two listeners, who are high-performers in the semitone tasks used in Chapter 2, to investigate the number of dimensions of sensitivity that enable tonal discrimination. Multidimensional scaling applied to the results yields 3 dimensions for each of two listeners. In chapter four, I take a much closer look at the interaction between judgments of major and minor and the strength with which the tonic is established in a sequence. We find that tonic strength interacts with the strength of major and minor cues in the tone-scramble, influencing the judgments of listeners beyond what the major and minor tones on their own would elicit.

Introduction

The major and minor modes in Western music have very different impacts on many listeners. The differences between them are not just striking, but the two modes consistently labeled as sounding happy and sad respectively. The large differences between the qualities evoked by these two modes get me interested in what other kinds of musical differences could create even partially analogous qualitative differences in listeners. It is the questions about what causes the differences in major and minor percepts, and other analogous qualities that fuel this dissertation.

I first question the role scale degree 6 plays in the percept of major or minor. Scale degree 3 and 6 both move down a half step when moving from major to minor, but only scale degree 3 is found in the major and minor chords based on the tonic of a scale. Is it the mode that matters? Or is it only the chord? I ask these questions in chapter 1.

Second, I question other percepts that may influence judgments that may be analogous to major and minor judgments. Scale degrees 3, 6, and to an extent 7, all move down a half step when moving from major to minor mode. However, some notes in the chromatic scale never show up in either of the major or minor diatonic scales. What about making judgments a half step apart that use those tones? Will they work the same way that the major and minor judgments do? I ask these questions in chapter 2.

With experiments using large groups of people, I am forced to make concessions as to how much I can ask of them. Ideally, I would like to do a large population study and ask them to compare every combination of notes, but I am forced to limit the experiment to what the participants can do in a reasonable amount of time. However, it may be much more practical to ask only a few people to compare every combination of notes. This is the approach I take in chapter 3.

In chapter 4, I turn my attention back to the original question of major and minor. Though it must be true that the tones that distinguish the major and minor modes and chords must be the cause of the major and minor feeling we get, that may not be the whole story. What tone is major and what tone is minor must be defined in terms of a tonic. While the same tonic can exist in both major and minor modes, it may be the case that the tonic still has an impact on the way major and minor is experienced. Maybe it is the case that the strength of the tonic allows the major or minor feeling to be heard more clearly. It is in this chapter that I ask what that impact might be of the tonic as it interacts with major and minor impacts on listeners.

Chapter 1

Scale Degrees Three and Six Influence Judgments of Major Versus Minor Mode Via the Same Auditory Mechanism

Abstract

This study investigated listeners' ability to classify (with feedback) tone-scrambles (rapid, randomly ordered sequences of pure tones) as major versus minor. In four different conditions, tone-scrambles differed in rate of presentation and/or the notes they included from the major versus minor diatonic scales. In the slow (fast) "thirds" condition, each tone-scramble contained 32 (64) tones presented in 2.08 sec; a major (minor) tone-scramble in this condition comprised equal numbers of G_5's, G_6's, D's and B's (B♭'s). A major (minor) tone-scramble in the corresponding, slow and fast "sixths" conditions comprised E's (E♭'s) instead of B's (B♭'s). The distribution of proportion correct across 120 listeners was strongly bimodal in each of the four conditions with distinct modes near chance and perfect performance. In addition, performance was strongly correlated across all four conditions; participants who performed well in one condition tended to perform well in all conditions. Participants (both high and low-performers) did slightly better in the "fast" than in the "slow" conditions and slightly better in the "sixths" than in the "thirds" conditions; however, these effects were dwarfed by the difference in performance between the two modes of the distribution. These findings suggest that (1) skill in these four tasks is

conferred by the same perceptual mechanism, and (2) this mechanism is accessible to listeners in the high-performing group but not to listeners in the low-performing group.

Introduction

The major and minor diatonic scales are central to western music. Many listeners experience music in the major mode as sounding happy, whereas music in the minor mode is said to sound sad, and indeed, substantial research confirms this impression ((Crowder, 1984, 1985 a, b; Gagnon & Peretz, 2003; Gerardi & Gerken, 1995; Heinlein, 1928 (as reanalyzed by Crowder, 1984); Hevner, 1935; Kastner & Crowder, 1991). However, other studies indicate that the distinction between major versus minor modes is less vivid to many listeners than its importance in western music might suggest. Halpern, 1984, and also Halpern, Bartlett & Dowling, 1998 found that listeners rated melodies that differed by mode as more similar than those that differed by contour or rhythm. They also found that subjects had more difficulty telling the difference between melodies that differed only in mode compared to those that differed in either contour or in rhythm. In addition, Blechner (1977) (replicated by Crowder (1985a)) presented participants with triads comprising a tonic, a dominant and a middle tone whose frequency was varied along the continuum between the minor and major third. The task was to judge whether the middle tone was closer to the minor or to the major third. Some of his participants achieved very sharp psychometric functions of frequency in this task; others, however, produced nearly flat psychometric functions suggesting that they were unable to hear much difference between the chords that included the major versus the minor third.

Blechner (1977) and Crowder (1985) both found participants who were able to make the distinction between major and minor, and those who were not, as well as few who performed in the middle. These results hint that the ability to discriminate major versus minor musical modes may conform to a bimodal distribution. Chubb et al. (2013) provide direct evidence that this is true. In this experiment, participants were asked to classify as major versus minor a new class of stimuli called "tone-scrambles." A tone-scramble is a rapid, random sequence of tones conforming to a specified histogram. The major (versus minor) tone-scrambles used by Chubb et al. (2013) contained 8 tones of each of the notes G_5, D_6, and G_6, and B_5 (or $B\flat_5$ in the minor condition). Participants strove to classify the tone-scramble presented on each trial as either major or minor, and received feedback after each trial. This task yielded a strikingly bimodal distribution in performance. Roughly 70% of listeners performed very poorly achieving proportions correct only slightly above chance; the other 30% achieved proportions correct near perfect.

The tone-scrambles in the experiments of Chubb et al. (2013) used only scale degrees 1,3,5 and 8. In particular, the majorness versus minorness of the stimuli used in this study was signaled exclusively by the presence of the major versus minor 3rd in the tone-scramble. However the major versus minor diatonic scales differ not only in the third degree of the scale. In the harmonic minor scale, degrees 3, 6 are flattened relative to degrees 3, 6 of the major scale. (In the natural minor scale, degrees 3, 6 and 7 are flattened.) This suggests that scale degree 6 is likely to be important in signaling the majorness versus minorness of the mode of a melody.

The bimodality of performance observed in the Chubb et al. (2013) study suggests that

1. there exists a perceptual mechanism M that is differentially activated by tone-scrambles that differ in containing major thirds versus minor thirds, and
2. listeners who perform well in the classification task used by Chubb et al. (2013) possess and can access M but listeners who perform poorly either do not possess M or cannot access it.

In the current study we ask: What factors determine sensitivity to differences in mode produced by varying the sixth degree of the scale? If sensitivity to differences in scale degree 6 is conferred exclusively by M, then we should find that performance in classifying tone-scrambles that differ only in scale degree 6 should mimic performance observed in classifying tone-scrambles that differ only in scale degree 3. Specifically, performance in classifying tone-scrambles that differ in only scale degree 6

1. should be strongly correlated with performance in classifying tone-scrambles that differ in only scale degree 3, and hence
2. should conform to a bimodal distribution similar to the bimodal distribution observed by Chubb et al. (2013).

There are, however, several reasons to think scale degree 6 might operate differently from scale degree 3 to mediate performance in a tone-scramble classification task. First, scale degree 6 is not used in the triad based around the tonic. A major chord in the tonic key uses scale degrees 1, 3 and 5. If the chord based around the tonic key is what influences people to make this distinction, it is possible that scale degree 6 will not work the same way. In addition, the melodic minor key flats the 6^{th} scale degree only in descending sequences, using the unflattened 6th in ascending sequences. This suggests that scale degree 6 may not be as effective a cue to majorness versus minorness as scale degree 3. It is also true that in the melodic and natural minor scales, the sixth scale degree is only a half step away from scale degree 5. If participants are especially sensitive to the presence of half step intervals in a tone-scramble, then they may be able to use this cue to classify tone-scrambles differing in degree 6; however, this cue will not help listeners to classify tone-scrambles differing in scale-degree 3.

Our primary focus is on how differences in scale degree 6 mediate performance in tone-scramble classification. In addition, however, we will investigate how performance in tone-scramble classification depends on stimulus speed by varying stimulus speed across two levels. In Chubb et al. (2013), participants who were not able to achieve a high percent correct in a major versus minor judgment using tone scrambles were sometimes able to achieve a high percent correct using chords. It may be the case that the inability to perform this task using tone-scrambles comes from the inability to integrate the whole stimulus across time. By using a faster stimulus, maybe the notes will blur together for the participant in such a way that it sounds more like a chord, enabling participants who cannot hear the difference in the slower tone-scrambles to hear the difference in the faster ones.

Methods

Stimuli

Each participant was tested in four, separately blocked, task conditions: a Slow-3 condition, a Slow-6 condition, a Fast-3 condition, and a Fast-6 condition. In the Slow-3 and Slow-6 conditions, each scramble consisted of 32 tones, each of which was 65ms in duration. In the Fast-3 and Fast-6 conditions each scramble consisted of 64 tones, each tone was 32.5 ms in duration. Thus, each tone-scramble in each of these four conditions lasted 2.08 seconds. All tones were pure tones drawn from the equal tempered scale and windowed using a raised cosine function:

$$W(t) = \begin{cases} \dfrac{1-\cos\left(\dfrac{\pi t}{k}\right)}{2} & 0 \le t < k \\ 1 & k \le t \le D-k \\ \dfrac{1-\cos\left(\dfrac{\pi (t-D)}{k}\right)}{2} & D-k < t \le D \end{cases} \quad (1)$$

where D = 65 ms and k = 22.5 ms (D = 32.5 ms and k = 11.25 ms) for the tones used in the Slow-3 and Slow-6 (Fast-3 and Fast-6) tone-scrambles.

All tone-scrambles in the Slow-3 and Slow-6 conditions contained 8 G_5's, D_6's, and G_6's. In addition, a major tone-scramble in the Slow-3 (Slow-6) condition contained 8 B_5's (E_5's), and a minor tone-scramble in the Slow-3 (Slow-6) condition contained 8 $B\flat_5$'s ($E\flat_5$'s). All tone-scrambles in the Fast-3 and Fast-6 conditions contained 16 G_5's, D_6's, and G_6's. In addition, a major tone-scramble in the Fast-3 (Fast-6) condition contained 16 B_5's (E_5's), and a minor tone-scramble in the Fast-3 (Fast-6) condition contained 16 $B\flat_5$'s ($E\flat_5$'s).

Participants

120 listeners participated. All were UC Irvine undergraduates with self-reported normal hearing. The experimental procedures were approved by the UC Irvine Institutional Review Board.

The Experiment

At the outset of the experiment, each participant was given an on-line survey which prompted him/her to supply his/her years of musical experience; if a participant indicated that he/she had one or more years of musical experience, he/she was prompted to supply the age at which he/she had begun studying music, his/her primary instrument, the hours per week he/she practiced and whether or not he/she had ever sung in a choir. Each participant was prompted to supply the hours per week that he/she listened to music, his/her primary language, and his/her dominant hand.

After taking the survey, each participant was tested in each of the Slow-3, Slow-6, Fast-3, and Fast-6 conditions (counterbalanced in order across participants using a Latin square design). The participant was told that he/she would be attempting to classify major versus minor tone-scrambles. He/she was told that the major tone-scrambles would sound “happy,” and that the minor scrambles would sound “sad.”

In each condition, prior to testing, the participant was presented with eight examples of tone-scrambles of the specific type used in that condition, alternating between major and minor. Prior to each major (minor) example tone-scramble, the participant and was told, via a message printed on the video monitor “The next scramble will be HAPPY (major)” (“The next scramble will be SAD (minor)”).

Following these eight example stimuli, the participant was tested in four blocks of trials in this same condition, each containing 40 trials (for a total of 160 trials). The stimulus on any given trial had a 50% chance of being major or minor, and the number of major and minor trials was not forced to be equal. Feedback was given visually after each trial; specifically, if the participant had responded correctly (incorrectly), the word “CORRECT” (“INCORRECT”) appeared on the screen. After each block the message “You have finished this block of trials. Hit enter when you are ready to continue to the next block” appeared on the screen. Participants were verbally encouraged to take breaks in between blocks to avoid fatigue and keep performance levels as high as possible.

Results

We treat the first 80 trials (the first two blocks) in each condition as practice and take the proportion correct from the last 80 trials (the last 2 blocks) as our dependent variable.

Histograms of the proportions correct achieved by our 120 participants in the Slow-3, Slow-6, Fast-3 and Fast-6 conditions are shown in Figure 1.1.

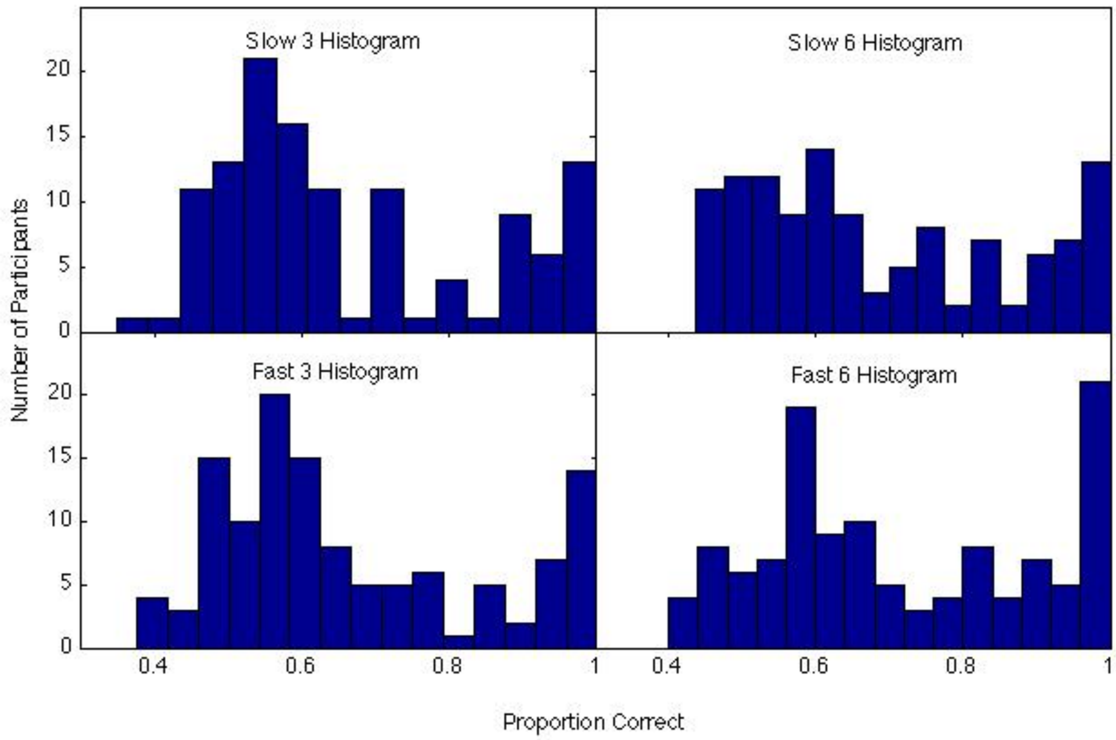

Figure 1.1: The distribution across participants of proportion correct in each of the four conditions. Each bar shows the number of participants in each range of scores. Note the bimodality of the histograms.

Performance in each of the four task conditions conforms to a bimodal histogram. One group, comprising roughly 60% of listeners, performs near chance; the other main group, comprising the other 40% of listeners performs nearly perfectly.

Table 1.1 gives the correlations in performance (as reflected by d') between different conditions. As is clear, all of these correlations are high suggesting that a participant's score on any one of these tasks is a good predictor of the participant's scores on the other three tasks.

Correlations Among Conditions				
	Slow 3	Slow 6	Fast 3	Fast 6
Slow 3	1.0000	0.8362	0.9127	0.8560
Slow 6	0.8362	1.0000	0.8704	0.8426
Fast 3	0.9127	0.8704	1.0000	0.8413
Fast 6	0.8560	0.8426	0.8413	1.0000

Table 1.1: Matrix giving the correlation between the d' values achieved by our participants in each pair of conditions.

Given that the population of listeners splits cleanly into high- versus low-performing groups, it makes sense to pose the following questions separately for each group:

1. Are listeners better at the major/minor classification task for 6^{th}s versus 3^{rd}s?
2. Are listeners better at the major/minor classification task for fast versus slow tone-scrambles?

We first extract two subgroups of participants, the high-performers and the low-performers. The high performers consist of those who scored a percent correct of 85% or higher across all four conditions (N = 28). The low-performers consists of those who scored 65% or lower across all conditions (N = 64). Participants in neither of these two groups are not included in the current analysis (N = 28).

We will perform a 2x2 repeated measures ANOVA, using the high- and low-performing group split as a between subject factor, where the two factors are stimulus speed and target-note.

Instead of percent correct, we use d' as our dependent measure in each of these two ANOVAs. There are two reasons to prefer d' in this context:

1. The percent correct achieved by a listener in a given condition depends both on the participant's sensitivity to the difference between the major versus minor tone-scrambles used in the condition and also on the criterion the participant uses in making his/her decisions. d' reflects the participant's sensitivity purified of performance degradations due to suboptimal selection of response criterion.
2. The proportion correct achieved by any given participant in a given condition must be viewed as a random variable X with expectation p equal to the true probability that this participant produces a correct response on any given trial in this condition. The variance of X is $p(1-p)$. If $p = 0.9$, this variance is 0.09. By contrast, if $p = 0.5$, this variance is 0.25. Thus, the variance of proportion correct depends strongly on p

violating the assumption of homogeneity of variance required for ANOVAs. Using *d'* fixes this problem.

The ANOVA yields a significant main effect of speed: $F(1,90) = 5.090$, $p = .026$, a main effect of note: $F(1,90) = 5.862$, $p = .017$, and there is no significant interaction between the two: $F(1,90) = 1.316$, $p = .254$. The interactions between speed and performance group ($F(1,90) = 2.086$, $p = .152$), note and performance group ($F(1,90) = .111$, $p = .739$), and speed, note, and performance group ($F(1,90) = .548$, $p = .461$) are not significant. Table 1.2 shows the results.

2x2 Repeated Measures ANOVA		
Effect	F(1,90)	p
Speed	5.090	0.026*
Note	5.862	0.017*
Speed*Note	1.316	0.254
Speed*Performance	2.086	0.152
Note*Performance	0.111	0.739
Speed*Note*Performance	0.548	0.461

Table 1.2: The results of the 2x2 repeated measures ANOVA using note and speed as the two factors, and performance group as a between-subject factor and *d'* as the dependent measure. Only the main effects of speed and of note are significant.

Figure 1.2 shows the means for the different conditions. Dashed (solid) lines show the results for the high-performing (low-performing) group. Black lines show the mean d' values for the Slow-3 (on the left) and Fast-3 (on the right) conditions; white lines show the mean d' values for the Slow-6 (on the left) and Fast-6 (on the right) conditions. The main effect of note revealed by the ANOVA is reflected in this figure by the fact that the white solid line is above the black solid line, and the white dashed line is above the black dashed line. The main effect of speed is reflected by the fact that all four of the lines show an increase (some more strongly than others) from the slow condition to the fast condition. Although these effects are reliable (as indicated by the ANOVA results reported above), the effect sizes are small: The mean difference in d' between the "6" versus "3" condition and "fast" versus "slow" condition is 0.120 and 0.083.

By contrast, the mean difference in d' between the high- and low-performing groups is 3.035, more than an order of magnitude greater than the within-group effects due to variations in task conditions.

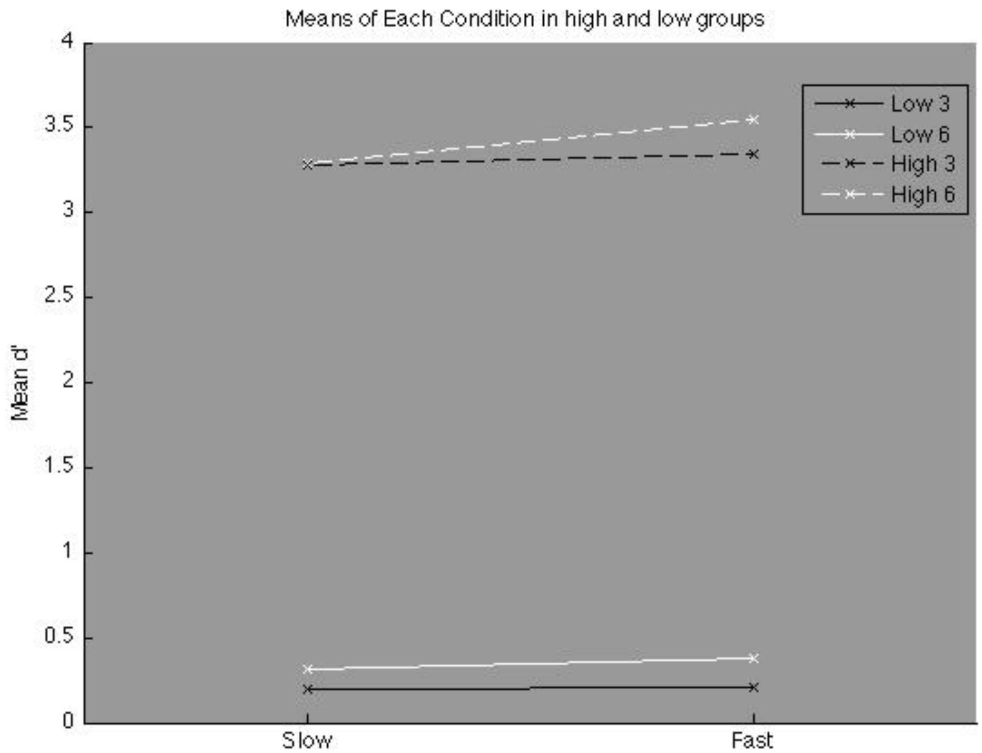

Figure 1.2: The means of the 4 different conditions for the low- and high-performing groups of participants. The ordinate gives the mean d' score obtained by each group in each condition. The abscissa shows the different speed conditions, the different colored lines represent different target note conditions, and the solid/dotted lines show the low- and high- performing groups respectively. Although the differences in d' due to target note (3 versus 6) and speed (slow versus fast) are significant, these effects (0.120 and 0.083 respectively) are dwarfed by the difference in d' between the low- and high-performing groups (3.035).

Correlates to the Tone-Scramble Task

Do any of survey responses predict performance in major versus minor tone-scramble classification? To investigate this question, we take as our dependent variable D' the value of d' achieved by each participant pooling across the last two blocks of trials in all four conditions. We then fit a regression using the survey responses as predictors of D'.

The variables used as predictors in the model were all derived from the responses to the questionnaire provided by participants. If a participant played an instrument that does not have discrete notes (violin, viola, trombone, and voice) his/her "Instrument" variable was assigned the value 1 and 0 otherwise. If he/she reported singing in a choir for a year or more, his/her "Choir" variable was assigned value 1, and 0 otherwise. If he/she spoke a tonal language as his/her first language, his/her "Language" variable was assigned value 1, and 0 otherwise. If he/she was left-handed, his/her "Handedness" variable was assigned the value 1, and 0 otherwise. In addition to these binary variables, the variables "Years of musical experience," "Hours practicing per week," and "Hours listening per week" were included as predictors.

Tables 1.3 and 1.4 give the numbers of participants who gave the different possible responses to the survey questions.

Survey Question Counts: Binary Questions		
	Yes	No
Participation in a Choir	30	90
Plays a Fretless Instrument	20	100

First Language is a Tonal Language	21	99
Left Handed	9	111

Table 1.3: The number of participants who answered yes and no to the binary survey questions.

Survey Question: Counts: Numerical Questions								
	0	1-5	6-10	11-15	16-20	21-25	26-30	31+
Years Experience	46	51	14	7	2			
Hours Practicing	111	7	1	1				
Hours Listening	1	41	27	20	11	7	4	9

Table 1.4: The numbers of participants who gave the different possible responses to the survey questions. All non-integer responses were rounded up to create this table.

The following predictors contribute significantly to the regression model: Years of Musical Experience ($t(112)= 4.012$, $p < .001$), Hours per Week Practicing ($t(112) = 2.302$, $p = .023$), and Choir ($t(112) = 2.362$, $p = .020$). The other predictors did not: for Hours per Week Listening, $t(112) = .099$, $p = .921$; for Instrument, $t(112) = -1.044$, $p = .299$; for Language, $t(112) = -.642$, $p = .522$; and for Handedness, $t(112) = 1.523$, $p = .131$. Although three of the predictors in the regression contribute significantly, together, they still only account for 24% of the variance: $SS_{Residual} = 151.565$, $SS_{Total} = 199.300$. The coefficients, standard error, t-values, and p-values for each variable can be found in Table 1.5.

Effect Size and Significance For Each Predictor				
Variable Name	Coefficient	Std. Error	t	p
Years of Experience	0.109	±0.053	4.012	<0.001*
Hours of Practice	0.150	±0.127	2.302	0.023*
Hours Listening	0.001	±0.014	0.099	0.921
Participation in a Choir	0.613	±0.510	2.362	0.020*
Plays a Fretless Instrument	-0.313	±0.588	-1.044	0.299
First Language is Tonal	-0.185	±0.566	-0.055	0.522
Left Handed	0.621	±0.800	1.523	0.131

Table 1.5: The effect size and significance for each predictor. Asterisks indicate significance at the 0.05 level.

One might wonder what proportion of the unexplained variance in the predictions of the regression model is due to the fact that the data observed for a given participant consists of binomial random variable and is thus inherently noisy. In order to investigate this, we conducted a simulation to estimate the variance due to measurement noise in our experiment. The simulation assumed that a given participant is characterized by a value of *D'* and a criterion *C* equal to the maximum likelihood estimates of these parameters derived from the actual data produced by this participant (pooled across the last two blocks in all four task conditions). We proceeded to perform 1000 simulations of the experiment. In each simulated experiment, each of the 120 participants performed 320 (simulated) trials (as in the actual experiment) with the odds of a hit, miss, correct rejection, and false alarm determined by the values of *D'* and *C* estimated for that participant. For each participant, a value of D' is estimated from each of the 1000 simulated data sets. The variance of this distribution of simulated values of D' provides an estimate of residual variance due to measurement error for that participant.

The sum (across our 120 participants) of the variances of D' estimated from the simulation was approximately 3.1, accounting for approximately 1.6% of the total variance (2% of the residual variance). We conclude that nearly all of the variance unaccounted for by the regression model derives from other sources than measurement error.

Learning Across Blocks

Do our data provide any evidence of improvement in performance in major versus minor tone-scramble classification due to learning? Any participant whose average proportion correct across all blocks in all four conditions was 0.5 cannot possibly have demonstrated any net improvement. The same is true of any participant who performed perfectly across all blocks in all four conditions. Accordingly, we restrict our analysis to include only those participants who achieved proportions correct between 60% and 90% across all blocks in all four conditions. There were 46 participants in this subset of potential learners.

We first investigate for each of the four conditions separately, whether learning occurred in that condition. For each participant in our subset of potential learners, we first estimate d' for each of the four blocks in the given condition. We then fit a line to these four d' values. For a given participant k, the slope S of this line was taken as a measure of learning in that condition. The mean S value for each condition was as follows: for the Slow-3 condition, .456; for the Slow-6 condition, .445; for the Fast-3 condition, .434; for the Fast-6 condition, .499. Then a t test was performed to test the null hypothesis that the mean value of S was less than or equal to 0. This null hypothesis is rejected decisively for all four conditions: for the Slow-3 condition, $t(45)= 8.97$, $p < .001$; for the Slow-6 condition, $t(45) = 10.735$, $p < .001$ for the Fast-3 condition, $t(45) = 9.743$, $p < .001$, and for the Fast-6 condition, $t(45) = 11.460$, $p < .001$.

The high correlations in performance observed across the four task conditions suggest that these tasks all engage the same auditory subsystem. This suggests that learning may transfer across the different tasks. We investigate this issue as follows. For each participant in our subset of potential learners, we first estimate d' for each of the 16 blocks in the order in which the participant performed the experiment. We then fit a line to these 16 d' values. For a given participant k, the slope S_{all} of this line was taken as a measure of learning. Then a t test was performed to test the null hypothesis that the mean value of S_{all} was less than or equal to 0. The null hypothesis is rejected, $t(45) = 11.740$, $p < .001$. Figure 1.3 shows the mean d' for all of the participants in each block minus the mean of all 16 mean d' scores. The error bars are 95% confidence intervals. The separate lines indicate where the participants switched conditions. Participants improved the most in the first four blocks, and seem to level off after the first condition, possibly with small improvement throughout the remainder of the experiment.

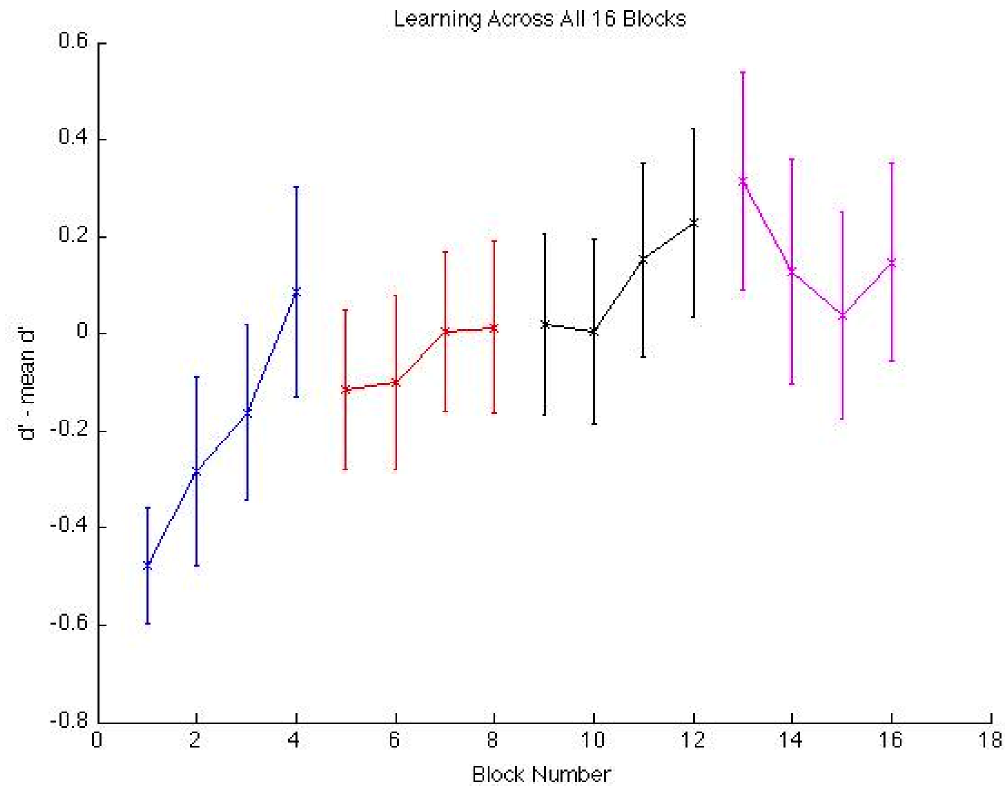

Figure 1.3: Performance across the entire experiment with respect to block number. The abscissa shows block number in the order that each participant performed the task. The ordinate shows the mean d' of the participants at that block minus the mean d' of the whole experiment. Note that learning occurs quickly in the beginning and levels out toward the end of the experiment.

Discussion

The most striking aspect of the data is that for all four conditions, participants show a bimodal distribution in their ability to perform the major-versus-minor classification task. In addition, performance is strongly correlated between all four task variants. This suggests that performance in the '6' tasks (Slow-6 and Fast-6) is mediated predominantly by the same mechanism as performance in the '3' tasks. The finding that performance is slightly but significantly better in the '6' tasks than in the '3' tasks may reflect a secondary cue available in the '6' tasks but not in the '3' tasks. However, whatever mechanism is responsible for the bimodal distribution of performance in the target-note =3 condition is responsible for performance in the target-note=6 condition as well.

While many participants perform better in the target-note=6 conditions than they do in the target-note=3 conditions, the advantage in performance conferred by this factor is very small in comparison to the massive difference in performance between the high- versus the low-performing groups. More importantly, participants who are in the low group for one of the tasks tend to be in the low group for all of the tasks. The participants in the low-performing group tend to perform near chance in all conditions; in particular, the average d' achieved by a participant in the low-performing group in the condition with the highest mean d' was .375, which corresponds to a percent correct of 57.44 if responses are made using an optimally placed criterion.

Correlates of skill in major versus minor tone-scramble classification

Of the different survey questions that the participants were asked, only three of the questions were useful for predicting the participants' D'. Interestingly, "Tonal first language" was not a useful predictor of success at this task, nor was "Hours per week listening to music."

The variables that emerged as useful predictors of performance in the tone-scramble task all reflect active engagement in music-production. These were: “Years of musical experience,” “Hours per week practicing,” and “Participation in a choir”. By contrast, passive exposure to music seems to have little relation to skill in the major-versus-minor classification task.

Although all three of “Years of musical experience,” “Hours per week practicing,” and “Participation in a choir” correlate significantly with performance in the tone-scramble task, the correlations between each of these predictors and d' are not large. With all of the predictors together, we still can only explain about a quarter of the variance. The primary source of skill in this task seems to reside outside the scope of these predictors.

The bimodal histograms observed in all four variants of the tone-scramble classification task suggest that (1) there exists an auditory mechanism M that responds differently to major versus minor tone-scrambles, and (2) the primary factor determining skill in (all four variants of) the tone-scramble task is possession of and access to M; high-performing participants have access to M whereas low-performing participants do not. We conclude that none of the variables included as predictors in the regression analysis correlate very strongly with whether or not a participant has access to M. A central question for future investigation is thus, what factors determine whether a participant has access to mechanism M?

Summary

This study used tone-scrambles (rapid, random sequences of tones) to investigate the mechanisms regarding sensitivity to major and minor. 120 listeners were tested in 4 different classification tasks. In each task, the listener strove (with trial-by-trial feedback) to classify the stimulus on each trial as major or minor, depending whether it contained sets of notes from the major or minor diatonic G scales. In each of the tasks, al tone-scrambles contained 8 G_5's, 8 G_6's, 8 D's to establish G as the tonic and D as the dominant of all tone-scrambles. In addition, major tone-scrambles in the "3" and "6" conditions contained 8 occurrences of either B♭ or E♭ respectively, and major tone-scrambles contained 8 occurrences of the tone a semitone higher.

Participants were tested in 4 conditions: slow-3, slow-6, fast-3, and fast-6. Though there was a significant difference between the slow and fast conditions, as well as between the 3 and 6 conditions, the effect sizes of the differences between the conditions were very small. Performance varied widely across listeners, but did conform to a bimodal distribution, such that the low group achieved d' scores of about 0.25, while the high group obtained d' scores of about 3.25. Correlations between all four of the conditions were very high. Because correlations were so high, and the bimodal distributions were so striking for all conditions, we conclude that those performers in the high group have access to a single mechanism that is primarily responsible for making the correct judgment all conditions, and that those in the low-performing group do not have access to this mechanism.

Survey questions revealed a significant effect of musical experience, hours of practice, and being in a choir for a year or more, but not for hours of music listened to, or playing an instrument with continuous notes, handedness, or initially speaking a tonal language. Though learning occurs throughout this task, initial learning seems to be very quick, and then levels out after about 200 trials.

Chapter 2

Semitone Differences in Pitch Judgments Using a Single Mechanism

Abstract

This study investigated the ability for a general population of undergraduates to distinguish between global statistics of different sequences of tones. The stimuli consisted of tone scrambles: quickly presented, randomly ordered sequences of pure tones, drawn from the equal-tempered chromatic scale. All tone scrambles used contained 32 tones. There were 6 conditions, five semitone conditions and a pitch height condition. Each semitone condition consisted of 8 copies of the intended low octave tonic, dominant, and high octave tonic, G_5, D_6, and G_6 respectively, as well as 8 copies of a note that the participant is meant to identify. We name the five semitone conditions the "2" condition, "3" condition, "4" condition, "6" condition, and "7" condition. In those conditions, the participant is meant to distinguish between A♭ and A, B♭ and B, C and D♭, E♭ and E, and F and G♭ respectively. In the pitch height condition, the participant is meant to distinguish between two distributions of notes that contain a large number of low notes vs. a large number of high notes, all of which span the same single octave as the other conditions. We find that participants tended to do best in the pitch height condition, followed by the "2", "3", and "6" conditions, which were all about the same. Participants performed more poorly on the "4" condition, and most poorly on the "7" condition. Though performance varied in their abilities at each task, most of the variance in the data for the semitone conditions can

be explained by a single mechanism model. Though this model does not account for everything in the data, it does explain most of the structure in our data set for the semitone task.

Introduction

Music is capable of eliciting many colors and emotions in humans. Questions of how exactly we perceive music and musical stimuli, however, remain largely unanswered. A basic supposition of the current work is that two musical stimuli can differ in quality for a listener only if they produce different levels of activation in one or more "mechanisms" in the human auditory system. We think of a mechanism as a collection of neurons capable of activating differentially when given different tonal stimuli. We are interested in the number and kinds of different mechanisms the brain may have to perceive tonality. While we are not aware of work specifically on the number of musical mechanisms, there has substantial work attempting to place musical tones into a space (Krumhansl 1979, Krumhansl and Kessler 1982, Shepard 1982, Lerdahl 1988, Chew 2014). If we imagine the brain has several independent mechanisms for tonality perception, then the dimensionality of the space that the notes occupy should correspond to the number of dimensions found in the space. The distance from note to note should correspond to the degree of perceptual similarity, with closer notes being more similar, and less similar notes being further apart.

Many efforts have been made to place notes into a perceptual space. Krumhansl (1979) had participants listen to either chords or scales to get the sense of a key. They were then asked to rate the similarity of all combinations of two notes in the context of the key they heard previously. She then used multidimensional scaling to arrange the notes into a space based on their perceived similarities. She found that the notes formed a three-dimensional conical shape with the notes in the major triad (her experiment only used a major key) at the smallest ring in the cone, the other notes in the major key at the second largest ring, and the notes not in the major key along the widest ring.

Krumhansl and Kessler (1982) had participants rate the 'goodness of fit' of various different probe tones presented following melodies in different keys. From these ratings, they were able to find average ratings each key had to other keys, and use those similarity ratings to derive perceptual distances between keys. They modeled the 24 keys in a four dimensional space. If the first two dimensions were projected into a two dimensional space, the model would show the circle of fifths. The third and fourth dimensions projected into a two dimensional space would show all keys in the same mode, a major third apart, represented as the same point. The keys form a circle separated by keys a half step apart. This arrangement puts the keys sharing the most notes with each other adjacent to one another (those closest on the circle of fifths, and the relative major/minor key), as well as the parallel major/minor keys and the major third above for major keys and the key a major third below for minor keys. As an example, the keys close to C major are G major, F major, A minor, C minor and E minor.

Roger Shepard (1982) has proposed several models of pitch perception. Shepard focuses mostly on octave equivalence and pitch height, but also incorporates an equivalence to whole steps and half steps, as he holds that all steps in a diatonic scale must be perceived as equally distant. He proposes a model with a double helix wound around a helical cylinder. The most complex of the models he proposes has 5 dimensions. The first dimension in his model is pitch height, the second and third dimensions are the chroma circle (notes arranged in a circle, with the closest notes being a half step apart) and the fourth and fifth dimensions are the circle of fifths (notes arranged in a circle with the closest notes being a fifth apart).

Lerdahl (1988) proposed a non-geometric model attempting to unify pitch perception, chord perception, and key perception. Lerdahl's model describes pitch space as existing in a key, and any combination of notes existing in a hierarchy of steps away from one another. For Lerdahl, notes are perceptually close if one can move from the first note to the second using a small number steps. Notes exist in a hierarchy, such that all the notes in the chromatic scale are on the lowest level, the diatonic notes are in the second level, the tones in the major triad based around the tonic are on the third level, the tonic and the fifth scale degree are in the fourth level, and the tonic is alone in the fifth level. A step in this model, consists of a move to the note directly above or below the starting note, along the highest or second highest level the note is a part of. For example, the tonic is four steps away from the note a half step above it. The tonic can only move to the fifth scale degree along the 4^{th} level of the hierarchy. From there you can move to the third scale degree along the 3^{rd} level of the hierarchy, From there, you can move to the second scale degree, along the 2^{nd} level of the hierarchy. Finally, from there you can move to the note a half step above the tonic along the lowest level of the hierarchy. While Lerdahl does not conduct experiments of his own, he does explain Krumhansl's data, as the number of steps in his model from tone to tone correspond fairly well with the similarity judgments made by Krumhansl's listeners. In addition, his model is able to account for many aspects of music theory.

Chew (2014) also placed the notes into a perceptual space with the goal of constructing a key finding algorithm, although she also claims her space is a perceptual representation. She places the notes in a three-dimensional helix, with the notes going up in fifths as a quarter turn along the spiral, and, as a result, separated vertically by a major third. Her work is mostly theoretical, but she does cite examples that show that her model is consistent with Krumhansl's tone profile data. Her model is also similar to Shepard's double helical model, except it does not give any weight to pitch height. In addition, she notes that the algorithm that comes from her model finds keys in about the same amount of time it takes a person to correctly identify the key to a melody.

Nearly all of these models attempt to place musical notes into a space, with distance reflecting similarity of notes. The one exception is Lerdahl's model, and even his has a notion of distance with what the model calls "steps." Each of these models, with its number of dimensions makes a prediction about the number of mechanisms to which the brain has access for evaluating tonal sequences.

The data supporting all of these models relies predominantly on judgments made by musicians or listeners with substantial musical experience. In addition, all of these data are derived using experimental paradigms akin to the probe-tone paradigm (Krumhansl and

Kessler, 1982) in which listeners rate the goodness with which various musical elements fit into various musical contexts. In paradigms of this sort, feedback is not appropriate because there is no right or wrong answer. Rather, the listener is the expert, and the data collected reflect the structure of the listener's expertise.

From our perspective, this approach has several shortcomings. First, we see no reason to limit the scope of our investigation to trained musicians. The mechanisms mediating musical experience should be present (in greater or lesser degrees) in all listeners. Second, feedback-free paradigms such as the probe-tone paradigm cannot measure the sensitivity of a listener to a given sort of stimulus variation. To measure such sensitivity, it is imperative to use a paradigm in which there is a right answer on each trial. To assess the sensitivity of the listener to the stimulus variation tested in a given condition, it is essential that the listener receive trial-by-trial feedback in order to optimize his or her response strategy.

An approach of this sort was taken by Chubb et al. (2013). This study analyzed the abilities of listeners to classify various sorts of musical stimuli as major vs. minor. Especially interesting results were obtained using tone-scrambles, randomly ordered sequences of pure tones presented at the rate of approximately 15 per second. In a basic condition from this study, every stimulus was a tone-scramble that contained 8 tones each with note G_5 (low tonics), G_6 (high tonics), and D_6 (dominants) and in addition either 8 tones with note B_5 (major thirds) or else 8 tones with note B-flat$_5$ (minor thirds). The task of the listener was to classify the stimulus as major (if it contained B_5's) vs. minor (if it contained B-

$flat_5$'s). The distribution of proportion correct was strikingly bimodal, with one mode (subsuming the majority, approximately 65%, of listeners) very close to chance performance, and the other very close to perfect performance.

The tone-scramble task used in the study of Chubb et al. (2013) partitions listeners very cleanly into two classes: High-performers vs. low-performers. However, this study left many important questions unanswered concerning the nature of the sensitivity revealed by the tone-scramble classification task. Crucially, the tone-scrambles used by Chubb et al. (2013) differed only in the third scale degree. We know that high-performers are sensitive to the difference between these two types of tone-scrambles and low-performers are not. What we do not know is how this sensitivity generalizes to other musical tasks.

The purpose of the current study is to address this question focusing specifically on variations of the tone-scramble task. Like scale degree 3, scale degree 6 also exerts a powerful influence on the perceived majorness vs. minorness of a melody. On the other hand, scale degree 2 exerts very little influence on majorness vs. innerness. Thus, for example, if the sensitivity that differentiates high- from low-performers is specific to major vs. minor, then we might expect performance in the tone-scramble task used by Chubb et al. (2013) (1) to correlate strongly with performance in a variant of the tone-scramble task in which stimuli differ in containing major vs. minor 6^{th}'s but (2) to correlate weakly with performance in a variant in which stimuli differ in containing notes of scale degree 2 vs. flatted scale degree 2. The experiment described below is designed to address questions of this sort.

Methods

Stimuli

Each participant was tested in 6 separately blocked conditions: a “2” condition, a “3” condition, a “4” condition, a “6” condition, a “7” condition and a “pitch height” condition. In each condition, stimuli consisted of 32 randomly ordered tones, each of which was 65 mms in duration with the entire tone-scramble lasting 2.08 seconds. All tones were pure tones drawn from the equal tempered scale and windowed using a raised cosine function:

$$W(t) = \begin{cases} \dfrac{1-\cos\left(\dfrac{\pi t}{22.5ms}\right)}{2} & 0 \le t < 22.5ms \\ 1 & k \le t \le 42.5ms \\ \dfrac{1-\cos\left(\dfrac{\pi (t-65ms)}{22.5ms}\right)}{2} & 42.5 < t \le 65ms \end{cases} \qquad (1)$$

All tone-scrambles, except those in the pitch height condition, contained 8 G_5’s, D_6’s, and G_6’s, as well as 8 target tones. In the 2 condition, the target tone was either an $A\flat_5$ or an A_5, a minor second or major second above the low tonic (G_5). In the 3 condition, the target tone was either a $B\flat_5$ or a B_5, a minor third or major third above the tonic. In the 4 conditions, the target tone was either a C_6 or a $D\flat_6$, a perfect fourth or a tritone above the tonic. In the 6 condition, the target tone was either an $E\flat_6$ or an E_6, a minor sixth or a major sixth above the tonic. In the 7 condition, the target tone was an F_6 or a $G\flat_6$, a minor seventh or major seventh above the tonic. Tone-scrambles in the pitch height condition contained either 5 G_5's, 4 $A\flat_5$'s, 4 A_5's, 4 $B\flat_5$'s, 4 B_5's, 4 C_6's, 1 $D\flat_6$, 1 D_6, 1 $E\flat_6$, 1 E_6, 1 F_6, 1 $G\flat_6$, and 1 G_6; or 1 G_5, 1 $A\flat_5$, 1 A_5, 1 $B\flat_5$, 1 B_5, 1 C_6, 4 $D\flat_6$'s, 4 D_6's, 4 $E\flat_6$'s, 4 E_6's, 4 F_6's, 4 $G\flat_6$'s, and 5 G_6's. For ease of viewing, the histograms of the two different types of tone-scrambles for each condition are displayed in Table 2.1 below.

	G5	A♭	A	B♭	B	C	D♭	D	E♭	E	F	G♭	G6
2 T. 1	8	8	0	0	0	0	0	8	0	0	0	0	8
2 T. 2	8	0	8	0	0	0	0	8	0	0	0	0	8
3 T. 1	8	0	0	8	0	0	0	8	0	0	0	0	8
3 T. 2	8	0	0	0	8	0	0	8	0	0	0	0	8
4 T. 1	8	0	0	0	0	8	0	8	0	0	0	0	8
4 T. 2	8	0	0	0	0	0	8	8	0	0	0	0	8
6 T. 1	8	0	0	0	0	0	0	8	8	0	0	0	8
6 T. 2	8	0	0	0	0	0	0	8	0	8	0	0	8
7 T. 1	8	0	0	0	0	0	0	8	0	0	8	0	8
7 T. 2	8	0	0	0	0	0	0	8	0	0	0	8	8

PH T. 1	5	4	4	4	4	4	1	1	1	1	1	1	1
PH T. 2	1	1	1	1	1	1	1	4	4	4	4	4	5

Table 2.1: The number of iterations of each tone in each type in each condition. The first two rows contain the histograms for the type 1 and type 2 scrambles respectively in the scale degree 2 condition. Every pair of rows thereafter contains the type 1 and type 2 scrambles for each condition respectively. The 6 conditions shown are scale degree 2, scale degree 3, scale degree 4, scale degree 6, scale degree 7, and pitch height.

Participants

140 listeners participated. However, the data from one participant was discarded because this participant issued a single, fixed response across all trials in one experimental condition. All were UC Irvine undergraduates with self-reported normal hearing. The experimental procedures were approved by the UC Irvine Institutional Review Board.

The Experiment

At the outset of the experiment, each participant was given an on-line survey which prompted him/her to supply his/her years of musical experience; if a participant indicated that he/she had one or more years of musical experience, he/she was prompted to supply the age at which he/she had begun studying music, his/her primary instrument, the hours per week he/she practiced. Each participant was prompted to supply whether or not he/she had ever sung in a choir the hours per week that he/she listened to music, his/her primary language, and his/her dominant hand.

After taking the survey, each participant was tested in each of the 2, 3, 4, 6, 7, and pitch height conditions (counterbalanced in order across participants using a Latin square design). The participant was told that he/she would be attempting to classify type 1 versus type 2 tone-scrambles.

In each condition, prior to testing, the participant was presented with eight examples of tone-scrambles of the specific type used in that condition, alternating between type 1 and type 2. Prior to each example type 1 (type 2) tone-scramble, the participant and was told, via a message printed on the video monitor "The next scramble will be TYPE 1" ("The next scramble will be TYPE 2").

Following these eight example stimuli, the participant was tested in two blocks of trials in this same condition, each containing 50 trials (for a total of 100 trials). The type of the stimulus on any given trial was equally likely to be 1 or 2; each participant received an equal number of type 1 and type 2 trials. Feedback was given visually after each trial; specifically, if the participant had responded correctly (incorrectly), the word "CORRECT" ("INCORRECT") appeared on the screen. After each block the message "You have finished this block of trials. Hit enter when you are ready to continue to the next block" appeared on the screen. Participants were verbally encouraged to take breaks in between blocks to avoid fatigue and keep performance levels as high as possible.

Results

For each participant in each condition, we treat the first 50 trials as practice and take the d′ value achieved across the last 50 trials as our dependent measure.

A central question motivating the current experiment is: How many dimensions of musical sensitivity are engaged by the different variants of the tone-scramble task tested here? To frame the analysis, it will be useful to consider a basic model in which there exists only a single such dimension.

A simple, 1-dimensional model. Suppose there exists in the human auditory system only a single mechanism M that is differentially activated by any of the tone-scrambles used in this study. More specifically, suppose that the activation produced by a given tone-scramble τ in this mechanism for a given listener k is

$$M_k(\tau) = S_k \sum_{j=1}^{32} f(\tau(j)) + \varepsilon \qquad (2)$$

where (1) $\tau(j)$ is the note of the j^{th} tone in τ, (2) S_k is the sensitivity of mechanism M as instantiated in participant k, (3) ε is a standard normal random variable, and (4) for any of the notes n in the equal-tempered chromatic scale from G_5 to G_6, $f(n)$ gives the relative activation produced in the mechanism M by a tone with note n, where the term "relative activation" is intended to signal that across the 13 notes n = G_5, A♭,..., G_6, the sum of $f(n)$ is equal to zero and the sum of $f^2(n)$ is equal to 1.

Under this model, in task T, we assume that participant k computes $M_k(\tau)$ for the tone tone-scramble τ presented on a given trial and judges the stimulus to be of Type 1 if $M_k(\tau) > Crit_T$ (note that the subscript signals that the criterion used by the listener may vary from task to task). In this case, the value of d' achieved by listener k in task T will be

$$d'_{k,T} = S_k \Delta_T \tag{3}$$

where τ_T^1 is a scramble of type 1 in task T and τ_T^2 is a scramble of type 2, and

$$\Delta_T = \left|\sum_{j=1}^{32} f(\tau_T^2(j)) - \sum_{j=1}^{32} f(\tau_T^1(j))\right|. \tag{4}$$

Under this model, then, we see that it should be the case that $d'_{k,T}$ is a separable function of listener sensitivity S_k and stimulus discriminability Δ_T. That is, we have a single set of Δ_T values that is invariant across listeners, all of which are multiplied by the sensitivity S_k of any given listener to produce the d′ values for that listener in all tasks T.

It should also be noted, however, that although the single-mechanism model implies that $d'_{k,T}$ is a separable function of S_k and Δ_T (Eq. 3), the reverse is not true. It is easy to imagine scenarios in which that Eq. 3 holds even though there exist multiple mechanisms that are sensitive to the different tone-scrambles used in the tasks T. For example, suppose listeners all possess two mechanisms differentially activated by the tones used to create our tone-scrambles with relative activation functions f_1 and f_2. Then Eq. 3 might hold, for example, if

$$\Delta_T = \left(\Upsilon_1^2(T) + \Upsilon_2^2(T)\right)^{1/2} \tag{5}$$

for Υ_i defined for i = 1, 2, by

$$\Upsilon_i = \sum_{j=1}^{32} f_i(\tau_T^2(j)) - \sum_{j=1}^{32} f_i(\tau_T^1(j)). \tag{6}$$

Testing the separable model. Table 2.2 gives the correlations (across all listeners) between the values of d' achieved in each pair of tasks. The most salient feature of this table is that the pairwise correlation in performance between the pitch height task and any of the other tasks is much lower than any of the other correlations in the table. In each of the 2, 3, 4, 6, and 7 tasks, the tone-scrambles to be differentiated all contain 8 G_5's, D_6's, and G_6's; in addition, in each of these tasks, Type-1 tone-scrambles contain 8 copies of another note n, and Type-2 tone-scrambles contain 8 copies of the note a semitone higher than n. For this reason, we will refer to these tasks as "semitone tasks." The pitch height task uses tone-scrambles that differ from each other both more strongly and in a very different fashion than the tone-scrambles used in the semitone tasks.

That the pitch height task affords response strategies that are not available in the semitone tasks is suggested the fact that performance tends to be better in the pitch height task than in the semitone tasks. This is shown in Figure 2.1 which gives the mean d' values achieved in each task. Whereas all of the semitone tasks yield mean values of d' in the neighborhood of 1.0 (or lower), mean d' for the pitch height task is greater than 2.5.

Task	3	4	6	7	PH
2	0.81	0.65	0.77	0.60	0.28
3		0.73	0.82	0.61	0.22
4			0.70	0.63	0.16
6				0.66	0.30
7					0.21

Table 2.2. Correlations in performance (as reflected by d' values achieved by all listeners) for all pairs of tasks.

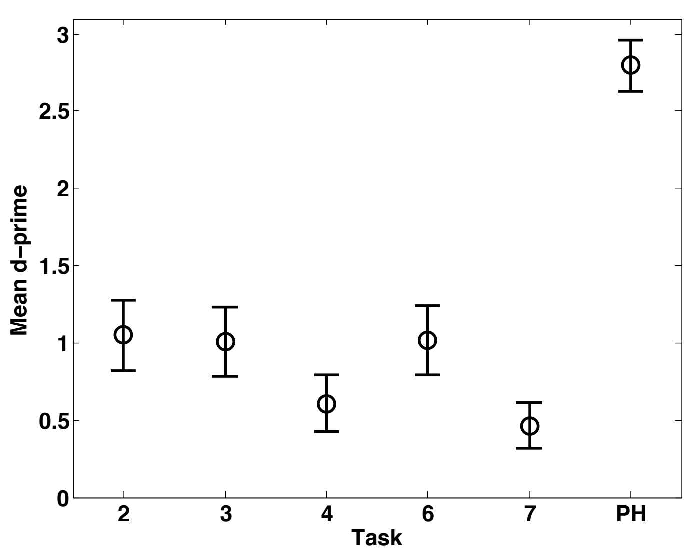

Figure 2.1: Mean values of d' achieved by all listeners in all tasks.

The low correlation between performance in the pitch height task and any of the semitone tasks suggests that the single-mechanism model is unlikely to hold for all six tasks. However, performance is highly correlated between all pairs of the different semitone tasks. This raises the possibility that a single mechanism may in fact mediate performance in all five of the semitone tasks. We address this question in the next section.

Testing the single-mechanism model across only the semitone tasks. We begin by deriving maximum likelihood fits of two models, a "full" and a "nested" model. Under the full model, listener k is assumed to respond "2" on a trial n in a given semitone task T if

$$d'_{k,T} + \varepsilon_{k,T,n} > C_{k,T}, \tag{7}$$

where $\varepsilon_{k,T,n}$ is a standard normal random variable, and $C_{k,T}$ is the criterion used by listener k in task T, and $d'_{k,T}$ reflects the sensitivity of listener k to the difference between the two tone-scramble stimuli to be classified in task T. In other words, under the full model, $d'_{k,T}$ is entirely unconstrained across conditions. Thus, the full model has 2*6*139 free parameters (two parameters for each listener in each task).

The nested model also assumes that responses are produced in accordance with Eq. (7). In addition, however, the nested model requires that the values of $d'_{k,T}$ satisfy Eq. (3) for some vector S_k of listener sensitivities and some vector Δ_T of stimulus discriminability values Δ_T

for different tasks T. Thus, the nested model has 6*139 (for the $Crit_{k,T}$'s) + 139 (for the S_k's) + 5 (for the Δ_T's) degrees of freedom.

To compare the nested and full models, we use a likelihood ratio test (e.g., Hoel, Port & Stone, 1971). This test makes use of the result (Wilks, 1944) that if the nested model captures the true state of the world, then the statistic $X = -2\log(\Lambda_{nested}/\Lambda_{full})$ is asymptotically distributed as chi-square with degrees of freedom equal the number of free parameters in the full model minus the number in the nested model.

The likelihood function for each of the full and nested models is

$$\Lambda = \prod_k \prod_T \Phi(C_{k,T})^{CR_{k,T}} \Phi(-C_{k,T})^{FA_{k,T}} \Phi(d'_{k,T} - C_{k,T})^{H_{k,T}} \Phi(C_{k,T} - d'_{k,T})^{M_{k,T}} \qquad (8)$$

where (1) the products range over all listeners k and all of the semitone tasks T, (2) Φ is the cumulative normal distribution function, (3) $CR_{k,T}$ ($FA_{k,T}$) gives the number of trials in which a stimulus of Type 1 was presented to listener k in task T and the listener responded correctly (incorrectly), and (4) $H_{k,T}$ ($M_{k,T}$) gives the number of trials in which a stimulus of Type 2 was presented to listener k in task T and the listener responded correctly (incorrectly). The symbols "*CR*," "*FA*," "*H*," and "*M*" reflect the standard concepts of "correct rejection," "false alarm," "hit," and "miss" from signal detection theory. Λ is implicitly a function of the model parameters. In the full model, these parameters are all of the criteria $C_{k,T}$ and all of the d-primes, $d'_{k,T}$. In the nested model, the d-primes are given by Eq. (3).

Note that Λ will be undefined if there exists any listener k for whom either $FA_{k,T}$ = 0 or $M_{k,T}$ = 0 in any task T. This condition was in fact satisfied by a number of high-performing listeners in a number of different tasks. To insure that Λ is well defined, we altered the data as follows. If $FA_{k,T}$ = 0 for a given listener k in a given task T, we doctored the data by setting $FA_{k,T}$ = 1 and incrementing $CR_{k,T}$ by 5. Similarly, If $M_{k,T}$ = 0 for a given listener k in a given task T, we doctored the data by setting $M_{k,T}$ = 1 and incrementing $H_{k,T}$ by 5. The choice to add 5 was largely arbitrary; varying this choice did not affect the results of our analysis in any important way.

When applied to the data for all listeners across the five semitone tasks, this test yields $\chi^2(df{:}\,551) = 1382,\ p \approx 0$, decisively rejecting the separable model. We conclude that the separable model (including the constraint of Eq. (3)) cannot account for the performance of our listeners. It also follows that the single-mechanism model is rejected.

Using the separable model to describe the data from the semitone tasks. Even though the separable model of performance (Eqs. (7) and (3)) is strongly rejected by the likelihood ratio test, it nonetheless provides a fairly good description of the data. This is shown by Figure 2.2 which plots $d'_{k,T}$ (estimated from the full model) for all five semitone tasks as a function of the value of d' estimated by taking $d' = S_k\Delta_T$ using the maximum likelihood estimates of S_k and Δ_T from the separable model. There is a strong relation between the values of d' predicted by the single-mechanism model and the unconstrained values of $d'_{k,T}$. In fact, the separable model accounts for 78% of the variance in the freely varying estimates of $d'_{k,T}$.

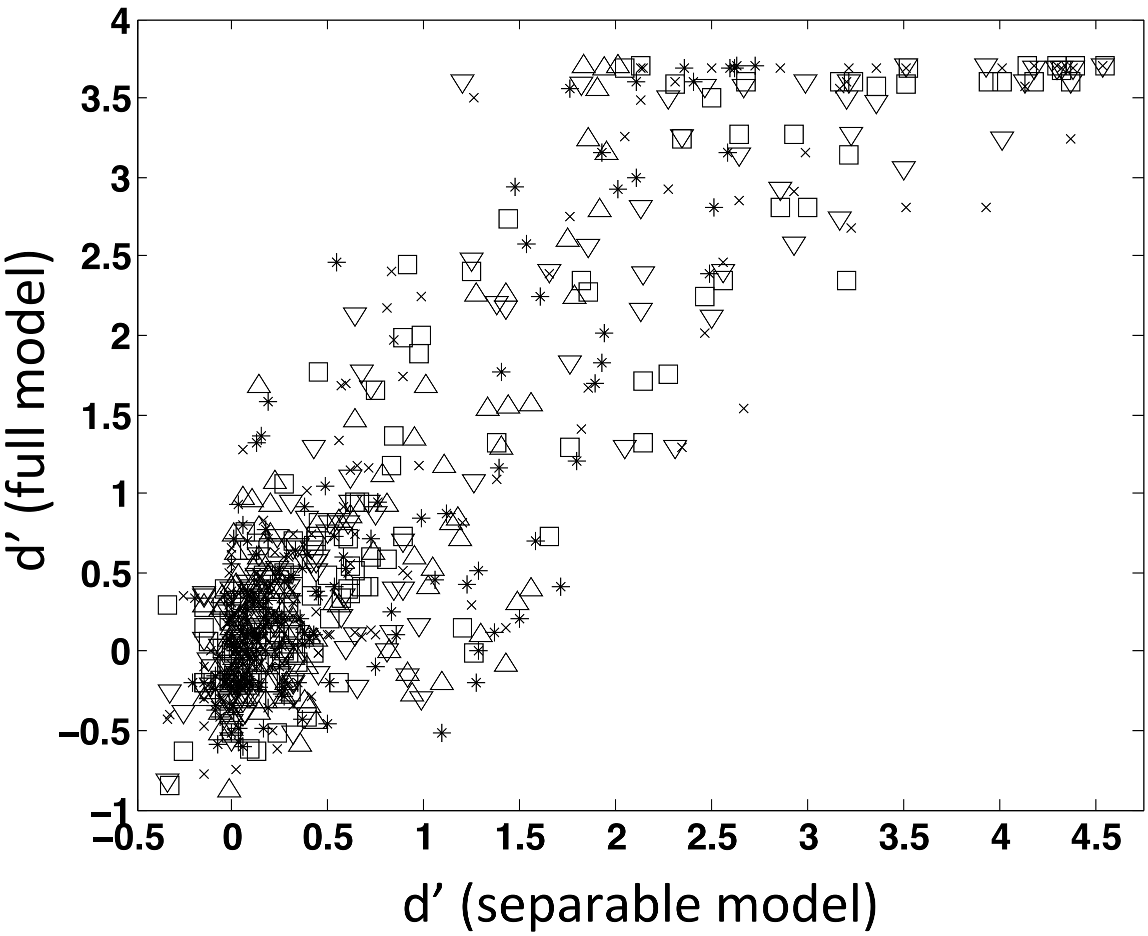

Figure 2.2. Scatter plot of the values of d′ estimated individually for each listener in each condition (ordinate) predicted by the estimates of d′ derived from the separable model (abscissa). Results for tasks 2, 3, 4, 6, and 7 are plotted respectively by x’s, squares, asterisks, downward pointing triangles, and upward-pointing triangles.

The separable model clearly captures a great deal of the structure in our data; indeed, even though the separable model is rejected by the likelihood ratio test, the residual deviations of the data from the predictions of the separable model show very little systematic structure. Let $d'_{full}(k, T)$ be the value of d' estimated by the full model for listener k in task

T, and let $d'_{sep}(k,T)$ be the value estimated from the separable model. Then the residual variation in d' unaccounted for by the separable model is given by

$$R(k,T) = d'_{full}(k,T) - d'_{sep}(k,T). \tag{9}$$

We can think of the rows of R as composing an ellipsoidal cloud of points in $\mathbb{R}^5$. What structure is discernible in this cloud? In particular, is it elongated or squashed in one or more directions? A simple way to address this question is to perform principal components analysis on the rows of R. We do this by applying singular value decomposition to the transpose of the matrix R. This yields a 5 by 5 orthonormal matrix C whose column vectors are the principal components of the elliptical cloud—i.e., the vectors corresponding to the principal axes of the ellipsoidal cloud, a diagonal matrix E of eigenvalues that reflect the degree to which the cloud is elongated in the directions of the corresponding principal components, and column-orthonormal 139 row by 5 column matrix L whose entries give the weights (or "loadings") of the different principal components in each of the row-vectors of R. The matrices C, E and L uniquely satisfy the condition that

$$R^T = CEL^T \tag{10}$$

where the superscript "T" indicates matrix transposition. The residual deviations of the data from the predictions of the separable model will be devoid of obvious structure if all of the eigenvalues (the diagonal entries of E) are equal, i.e., if the row vectors of R are distributed spherically around the origin of $\mathbb{R}^5$.

The resulting matrices C and E are

$$C = \begin{bmatrix} -0.6341 & 0.2189 & 0.4942 & 0.2854 & 0.4736 \\ 0.5988 & 0.5786 & -0.0900 & 0.3596 & 0.4114 \\ 0.4252 & -0.3959 & 0.5555 & -0.4173 & 0.4240 \\ -0.2400 & 0.2473 & -0.4868 & -0.6535 & 0.4661 \\ -0.0317 & -0.6320 & -0.4496 & 0.4338 & 0.4575 \end{bmatrix} \tag{11}$$

and

$$E = \begin{bmatrix} 9.3030 & 0 & 0 & 0 & 0 \\ 0 & 7.7005 & 0 & 0 & 0 \\ 0 & 0 & 7.4353 & 0 & 0 \\ 0 & 0 & 0 & 6.6777 & 0 \\ 0 & 0 & 0 & 0 & 1.8526 \end{bmatrix}. \tag{12}$$

Note that the first four eigenvalues are roughly equal (the corresponding axis lengths, obtained by taking square roots, range from 2.58 to 3.05). By contrast, the eigenvalue for the last principal component (the rightmost column of C) is substantially smaller (its corresponding axis length is 1.36). This signals immediately that there does indeed exist important structure in the cloud of points given by the row vectors of R.

What is the nature of this structure? The small eigenvalue E_5 (entry (5,5) of the matrix E) of the last principal component C_5 (the rightmost column of matrix C) signals that the residual error variance in the direction of C_5 is substantially smaller than the residual error variance in any direction orthogonal to C_5. Thus, the separable model has operated selectively to absorb error variance along this particular axis of $\mathbb{R}^5$. If the separable model were accounting for *all* structure in the data, the other four eigenvalues (reflecting the residual error orthogonal to C_5) should be approximately equal to E_5. This is clearly not the case. However, the near equality of the other four eigenvalues signals that the variations in R orthogonal to C_5 form a cloud that is approximately spherical, i.e., approximately devoid of elongation or compression along any axis through the origin of $\mathbb{R}^4$. Thus, although the residual variance orthogonal to C_5 is too large (in comparison to E_5) to be due to chance, neither does it display any clear structure for which we might attempt to account by elaborating our model. Finally, note that C_5 gives roughly equal weight to all five coordinate values. This tells us that the separable model has operated in an evenhanded fashion to account for performance in all five semitone tasks. This would not have been the case had we included the pitch height task in the analysis.

In light of the fact that the separable model seems to capture nearly all of the obvious structure in our data, it makes sense to take a close look at the patterns revealed by this model. This is the purpose of the next section.

What does the description provided by the separable model reveal? The maximum likelihood estimates of the stimulus discriminability values Δ_T for different tasks T are shown in Figure 2.3. Strikingly, Δ_2, Δ_3, and Δ_6 are nearly identical in value, all affording the highest level of performance achieved by our listeners. By contrast, each of Δ_4 and especially Δ_7 is substantially lower.

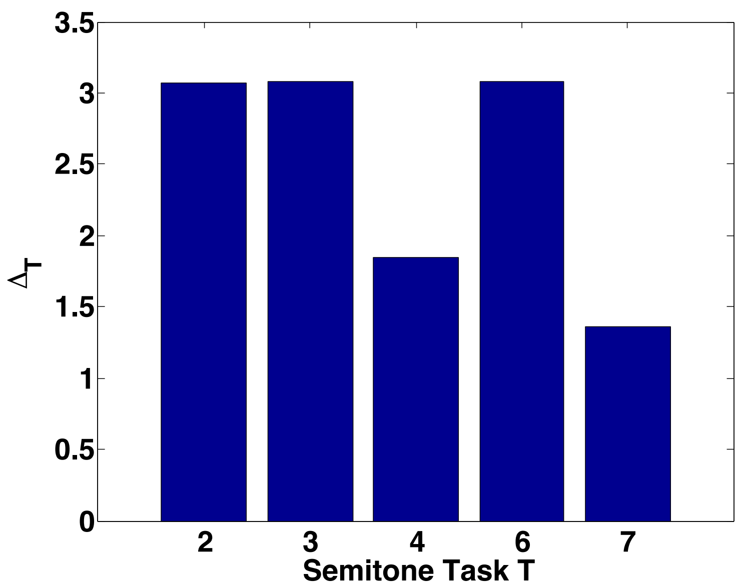

Figure 2.3. The maximum likelihood estimates of the stimulus discriminability values Δ_T in the separable model for the five different semitone tasks T.

How does the separable model predict that performance should be distributed across listeners? This question is answered most directly by focusing on the values of S_k yielded by the model. The histogram of estimated values of S_k is shown in the left panel of Figure 4. Note that this histogram highly skewed to the right; however, it does not appear bimodal. It should also be noted that the values of d' predicted for a given task T by the separable model can be obtained by multiplying the x-axis values in the left panel of Figure 2.4 by the value of Δ_T given in Figure 2.3. The right panel of Figure 2.4 shows the histogram of listeners predicted under the separable model to achieve different proportions correct in the 3 task. This plot is produced under the assumption that a given user listener uses an optimally placed criterion to produce responses. Note that, in contrast to the histogram in the left panel, this histogram appears strongly bimodal. Indeed, it is very similar to the histogram of proportion correct observed by Chubb et al. (2013) in the first experiment using the 3 task.

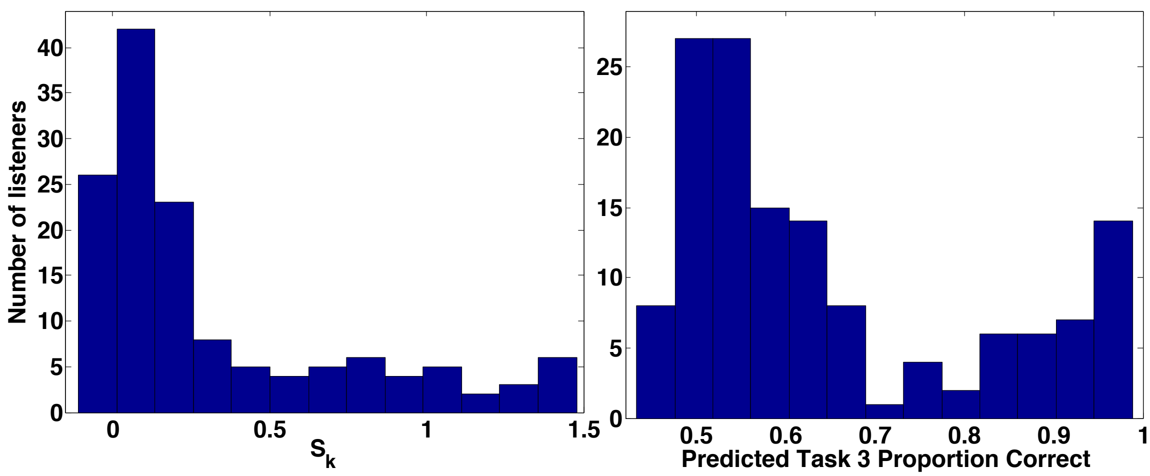

Figure 2.4. Histograms of predicted listener performance estimated from the separable model. The left panel gives a histogram of the values of S_k derived from the maximum likelihood fit of the separable model to the data. The right panel gives the histogram of listeners predicted under the separable model to achieve different proportions correct in the 3 task assuming each listener used an optimally placed criterion.

Discussion

The nature of the skill underlying performance in the semitone tasks. Chubb et al. (2013) documented a dramatic bimodal distribution in performance in classifying major vs. minor tone-scrambles of the same sort as were used in the 3 task in the current experiment. The high-performing listeners in the previous study evidently possessed some kind of auditory sensitivity that the low-performing listeners lacked. A central question motivating the current study was: what is the nature of this sensitivity?

The current results shed important light on this question. Although the 3 and 6 tasks (and possibly the 7 task) might be expected to engage a mechanism selective for major vs. minor, this is not true of the 2 and the 4 tasks. In each of these tasks one of the two target tones (A in the 2 task, C in the 4 task) occurs in both the G major and the G minor diatonic scales, and the other target tone (A♭ in the 2 task, D♭ in the 4 task) occurs in neither the scale. Nonetheless, performance in all five semitone tasks is strongly correlated (although the 2, 3 and 6 tasks are easier than the 4 and 7 tasks). This suggests that the same species of sensitivity gates performance in all of these tasks.

These observations suggest that the sensitivity that controls performance in all of the semitone tasks is not specific to major vs. minor but rather reflects a more general sensitivity to qualities that vary across music as the ensemble of notes present in the music is changed. We assume that the qualities used by our listeners to differentiate the various tone-scrambles used in these experiments depend on scale-degrees present in the tone-scramble relative to the tonic G. (It should be noted, however, that this assumption has not been tested; all of the stimuli we used contained a strongly established tonic of G. It is possible that a listener with absolute pitch, for example, might perform the various tasks in this study by listening for the presence of a single target tone in isolation without reference to the tonic G.)

Are the current results predicted by current models of tonality perception?

A prominent model of tonal sensitivity that has been successful in accounting for the ability of listeners to sense the changes in the key of music as it unfolds is the Spiral Array model of Chew (2014). In this section we investigate whether this model can provide traction in understanding the current results. In particular, can this model explain the differences in the stimulus discriminability values Δ_T observed for different semitone tasks T (as shown in Figure 2.3)?

In the spiral model of Chew (2014), notes are considered equivalent if they are separated by an integer number of octaves. Thus the equivalence class of all C's, for example, corresponds to a single node in the spatial structure that forms the basis of her model. She calls any such equivalence class a "pitch class." A given pitch class is assigned in Chew's (2014) model to a node on a helix. Nodes occur every 90 deg. along the helix, and the pitch class a quarter cycle upward along the helix from a given pitch class P is the pitch class a fifth above P (e.g., the G pitch class is a quarter turn up the helix from the C pitch class). Four steps along the circle of fifths from a given pitch class P are required to reach the pitch class a major third above P. Thus, in Chew's (2014) model the pitch class a major third above a given pitch class P is stacked vertically above P.

Chew (2014, p 21) says, "Starting from the pitch class spiral, the Spiral Array goes on to define points in the interior of the helix that represent higher level tonal constructs. Chords are represented as points that are the convex combination of their component pitches, a centroid of sorts." For purposes of assessing whether Chew's model can make sense of the current results, we will assume that a similar statement can be made about tone-scrambles, i.e., that the perceptual effect of a tone-scramble can be summarized in the context of Chew's model by a weighted sum of its component pitches.

Figure 2.5 shows how the Spiral Array model might be adapted to account for performance in the 3 Task, for example. The nodes along the spiral represent pitch classes. The large black disk represents the tonic (G) pitch class, the triangle represents the dominant (D) pitch class, the red disk represents the minor third (B♭) pitch class, and the green disk represents the major third (B) pitch class. In each of the tone-scrambles in this task (and in all of the other semitone tasks), the activation produced in the listener by each of the 16 G's is assigned to the location of the large black disk, and the activation produced by the 8 D's is assigned to the location of the large black triangle. In a Type 1 tone scramble, the activation produced by the eight B♭'s gets assigned to the

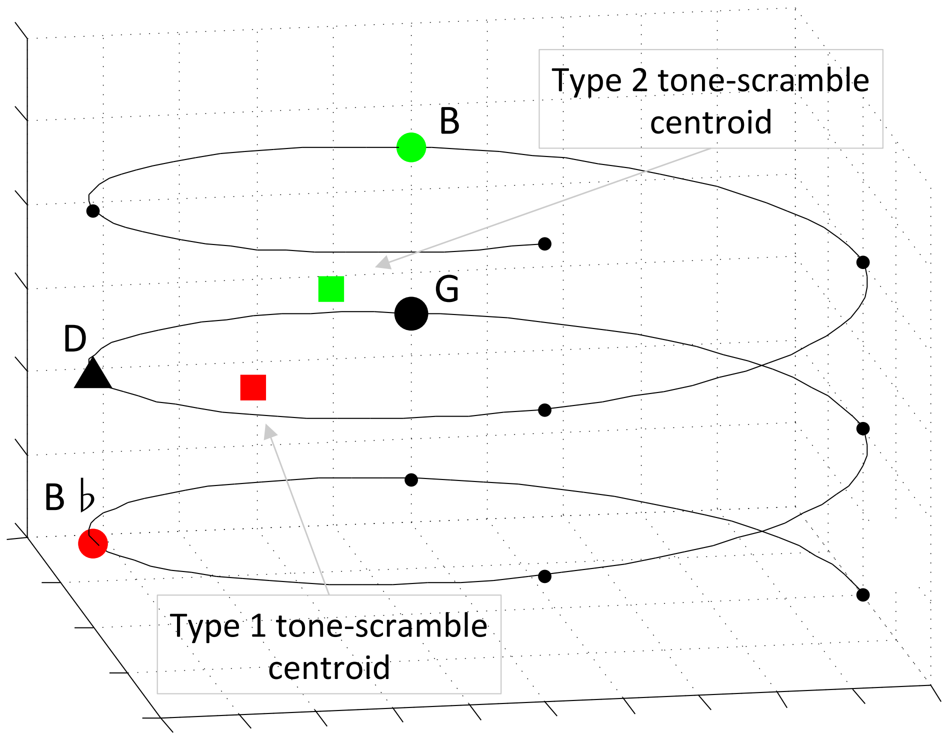

Figure 2.5: Using the Spiral Array model to account for performance in the 3 Task. The black nodes along the spiral (running from bottom to top) represent the pitch classes of A♭, E♭, B♭ (red disk), F, C, G (large black disk), D (large black triangle), A, E, B (green disk), G♭, and D♭. The red (green) square is the centroid of the activation produced in the Spiral Array by the 16 G's, 8 D's and 8 B♭'s (8 B's) that occur in a Type 1 (Type 2) tone-scramble in the 3 task under the assumption that tones of all pitch classes produce equal activation. We assume that the discriminability (Δ_T) of Type 1 vs Type 2 tone-scrambles in each semitone task T is proportional to the Euclidean distance separating the centroids of the activation produced in the Spiral Array by the two types of tone-scramble.

location of the red disk whereas in a Type 2 tone-scramble, the activation produced by the

8 B's gets assigned to the location of the green disk.

We assume that (1) the overall impact of a tone-scramble is summarized by the centroid of the activation in the Spiral Array produced by all the tones in the tone-scramble, and (2) in the 3 task, the discriminability of the Type 2 vs. Type 1 tone-scrambles (i.e., Δ_3) is proportional to the Euclidean distance between the centroids of activation produced by Type 1 vs Type 2 tone-scrambles (indicated in Figure 2.5 by the red and green squares respectively).

As yet unspecified are the levels of activation produced in the spiral array by tones from different pitch classes. If we assume that the weight exerted in the centroid by a given tone is proportional to the level of activation it produces, then these model parameters are critical in determining the values of Δ_T for different tasks T. For example, if the activation produced in the Spiral Array by B♭'s and B's is equal to that produced by G's and D's, then $\Delta_3 = 0.56$. By contrast, if the activation produce by B♭'s and B's is double (half) that produced by G's and D's, then $\Delta_3 = 0.90$ ($\Delta_3 = 0.32$).

We can obtain Δ_T's qualitatively similar to those observed in Figure 2.3 by assuming that the activations produced in the Spiral Array by different notes conform to Table 2.3. Thus, it is possible to account for the pattern of Δ_T's by assuming that in any task T, the activation produced by a tone from a given pitch class P conforms to Table 2.3.

Pitch class	G	A♭	A	B♭	B	C	D♭	D	E♭	E	F	G♭
Relative activation	2	5	5	5	5	2	2	2	5	5	1	1

Table 2.3. If the different pitch classes produce activations in the Spiral Array proportional to the corresponding relative activations, then $(\Delta_2, \Delta_3, \Delta_4, \Delta_6, \Delta_7) =$ (1.02 1.02 0.56 1.02 0.32). Note that this pattern is qualitatively similar to the pattern of the estimated values of Δ_T given in Figure 2.3.

Alternatively, one might imagine that the listener is able to use top-down attention to differentially adjust, for different tasks, the activation produced in the Spiral Array by tones from different pitch classes. In the 3 task, for example, perhaps the listener can deploy top-down attention in a way that makes B's and B♭'s produce 2.5 times the activation of G's and D's. In the 3 task, A's are never presented, but if one were to present an A in this context, one might find that it exerted very little influence on the listener's responses. Such a result would be consistent with the idea that the listener had deployed top-down attention in a way that heightened the activation produced in the Spiral Array by B's and B♭'s while suppressing the activation produced by A's.

We see then that it is possible to account for the stimulus discriminability values Δ_T observed in Figure 2.3 by appropriately selecting the activations produced in the Spiral Array by tones from different pitch classes; however, this merely replaces one mystery with another. Crucially, if all of target tones used in the different semitone tasks (A♭ and A (2 task), B♭ and B (3 task), C and D♭ (4 task), E♭ and E (6 task), F and G♭ (7 task)) produce the same levels of activation in the Spiral Array relative to G and D in the contexts of their individual tasks, then all of the Δ_T's will be equal. This holds true regardless of the activations produced in the Spiral Array by G and D. The Spiral Array model provides no reason why tones from these different pitch classes should behave differently from each other.

The relationship between tone-scramble tasks and the major and minor key profiles.

Krumhansl and Kessler (1982) asked each of 10 listeners (all of whom had had at least 5 years of formal musical instruction and had demonstrated substantial musical competence) to rate the goodness with which each of the 12 tones of the equal tempered chromatic scale fit into several different musical contexts. The four different musical elements used to establish the context of a major key were (1) the I chord (tonic, major third, dominant) of the scale, and the three chord sequences (2) IV-V-I, (3) II-V-I and (4) VI-V-I. Corresponding elements from a minor key were used to establish the minor context. The Major (Minor) profile is derived by averaging the ratings over four major (minor) scale contexts. The resulting Major and Minor tone profiles are shown in Figure 2.6.

The listeners in the study of Krumhansl and Kessler (1982) were making use of some neural apparatus to make their ratings. In this section, we address the following question: Is it this same apparatus that is required for successful performance in the semitone tasks in the current study?

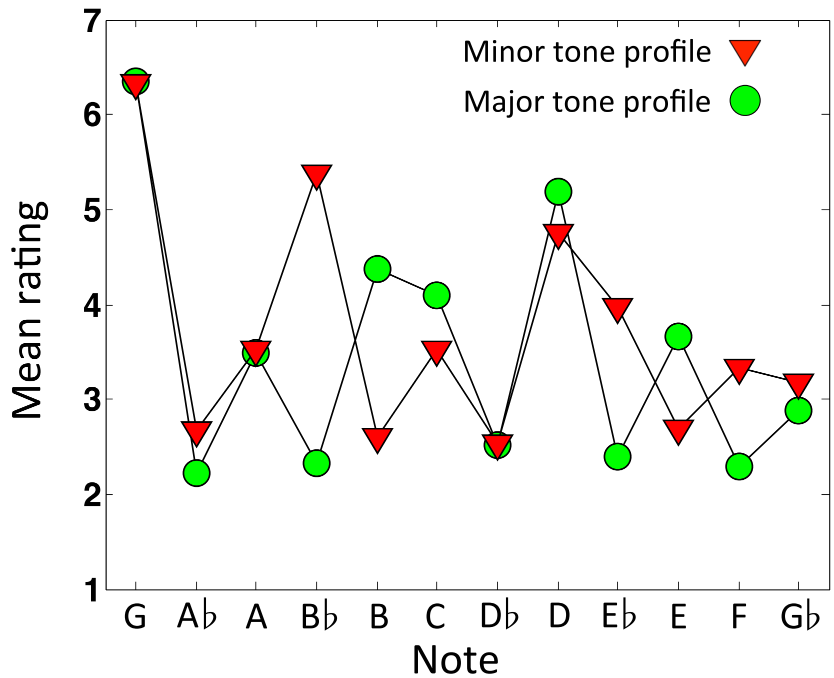

Figure 2.6. The Major and Minor tone profiles (Krumhansl & Kessler, 1982) plotted for the key of G used in the current experiments.

A given one of the semitone tasks requires the listener to discriminate two stimuli that can be viewed as sharing a context established by the 8 G_5's, 8 G_6's, and 8 D_6's that are common to all of the tone-scrambles in the semitone tasks. These context-defining tones, however, are common to both the G major and minor diatonic scales. Thus, they establish neither the Major nor the Minor context from which the tone profiles of Figure 2.6 are derived. Rather, they establish a neutral context in which we might expect especially notes of degree three, six and possibly also degree seven to introduce major or minor coloration. We assume that the discriminability conferred by this coloration is reflected by the difference between the mean rating given to major vs. minor thirds (the same goes for sixths and for sevenths) in either the major tone profile or the minor tone profile.

Note that the notes common to both the G major and minor scales (G, A, C, and D) do not vary much in the ratings they prompt in the major and minor profiles. Nor do the notes that occur in neither scale (A♭ and D♭). Accordingly, for purposes of the current analysis, we will assume that a given one of these notes activates the neural apparatus (used to produce the ratings reflected in Figure 2.6) to the same level (1) in each of the major and the minor contexts tested by Krumhansl & Kessler (1982) as well as (2) in the context established by the 8 G_5's, 8 G_6's, and 8 D_6's shared by the tone-scrambles in the semitone tasks.

The implication of these assumptions is that for any of the semitone tasks T in the current study, we predict the discriminability of the tone-scrambles in task T to be roughly in accordance with

$$\widehat{\Delta}_T \propto \frac{|Prof_{major}(n_2) - Prof_{major}(n_1)| + |Prof_{minor}(n_2) - Prof_{minor}(n_1)|}{2} \tag{13}$$

where $Prof_{major}$ and $Prof_{minor}$ are the major and minor tone profiles plotted in Figure 2.6, and n_1 and n_2 are the target notes that differentiate the tone-scrambles in task T (e.g., B♭ and B in the 3 task).

The predicted discriminabilities $\widehat{\Delta}_T$ are plotted by the black disks in Figure 2.7. The red triangles plot $|Prof_{minor}(n_2) - Prof_{minor}(n_1)|$ and the green circles plot $\left|Prof_{major}(n_2) - Prof_{major}(n_1)\right|$, and these are averaged to derive the open squares. The pattern of predicted discriminabilities in Figure 2.7 is similar to the pattern obtained in Figure 2.3 in several respects. First, $\widehat{\Delta}_7$ is predicted to be lower than the other four discriminabilities. Second, $\widehat{\Delta}_2$ is predicted to be similar in value to $\widehat{\Delta}_6$. The pattern in Figure 2.7 deviates from Figure 2.3, however, in other ways. In particular $\widehat{\Delta}_3$ is predicted to be substantially larger than all other $\widehat{\Delta}_T$'s whereas in Figure 2.3 we see that Δ_3 is equal to both of Δ_2 and Δ_6. Third, $\widehat{\Delta}_4$ is predicted to be roughly equal to $\widehat{\Delta}_6$ and $\widehat{\Delta}_2$. In Figure 2.3, Δ_4 is lower than Δ_2 and Δ_6.

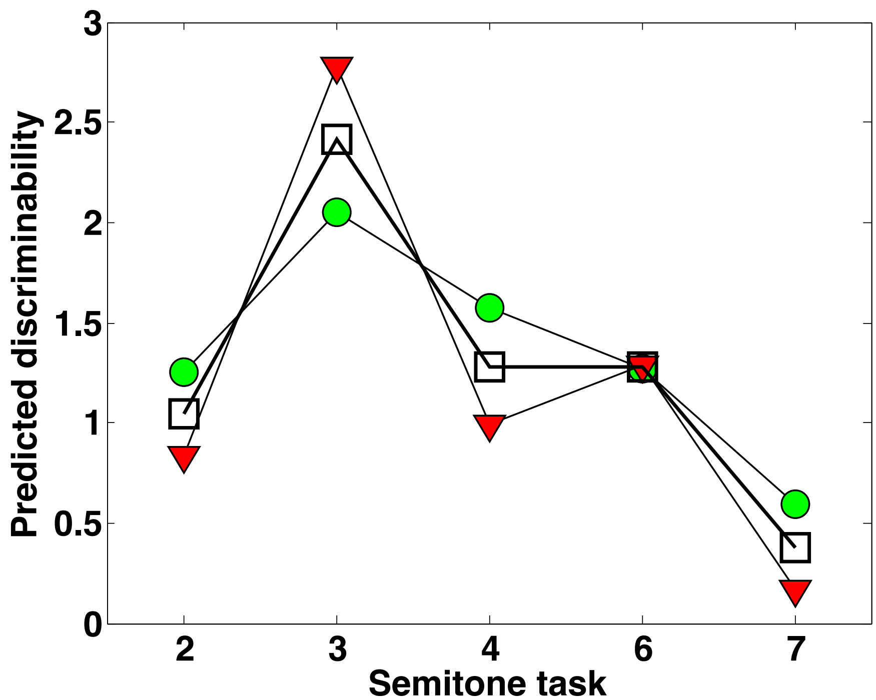

Figure 2.7. Semitone task discriminabilities $\widehat{\Delta}_{\mathrm{T}}$ predicted from the major and minor tone profiles (Krumhansl & Kessler, 1982) using Eq. (13). Red triangles (green circles) give predictions derived from the minor (major) tone profile. Open squares give the average of these two curves (the predictions of Eq. (13)).

Although Figure 2.7 diverges strongly in pattern from Figure 2.3, this is hardly surprising given the strong assumptions that we have had to make in order to derive these predictions. On the whole, Figure 2.7 suggests that the sensitivity deployed by the listeners in the study of Krumhansl & Kessler (1982) may well play an important role in enabling performance in the semitone tone-scramble tasks.

The difference between the predictions of Figure 2.7 vs. the empirical estimates obtained in Figure 2.3 may also be in part due to the fact that the listeners in the Krumhansl & Kessler (1982) study were trained musicians who brought to their judgments substantial knowledge of musical theory. It seems likely that some of these listeners may have been explicitly aware of the formal relationship of the note presented on a given trial to the context relative to which it was being assessed. For example, on a trial in which the probe tone presented was the third degree of the scale established by the chord sequence used to set the context, it seems likely that many of these listeners may have been explicitly aware of this fact. If so, their knowledge of the important role played in western music by scale degree three may have colored their ratings.

Figure 2.7 plots the predicted discriminabilities of the five semitone tasks under the assumptions that (1) each of the tones G, A, C, D, A♭ and D♭ produces activation that does not vary between major vs. minor contexts, and (2) the major vs. minor thirds, major vs. minor sixths and major vs. minor sevenths produce opposite activation in major vs. minor contexts. If we assume in addition that (3) the activations produced by G and D relative to the other tones match those in Figure 2.6, then we can we can predict the major and minor tone profiles from the semitone task discriminabilities plotted in Figure 2.3. These predicted major and minor tone profiles are shown in Figure 2.8.

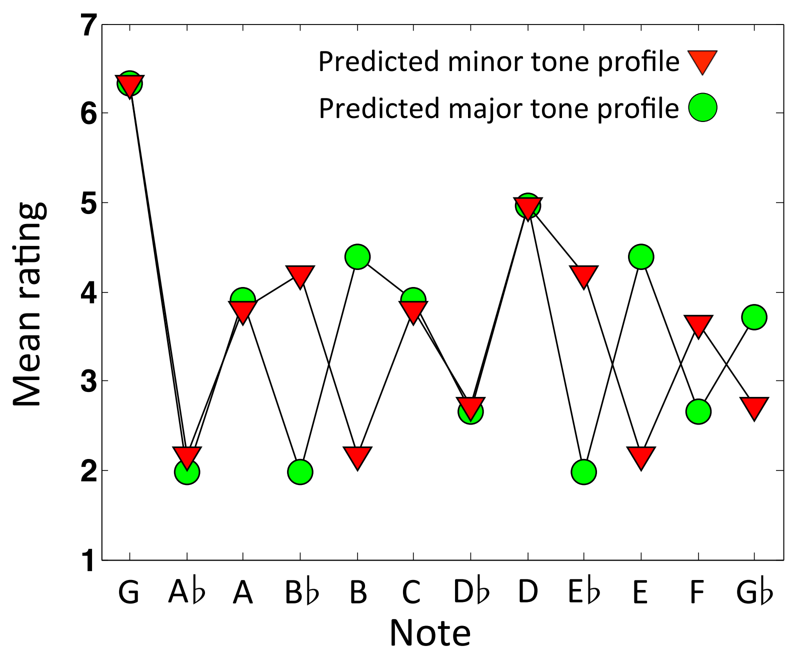

Figure 2.8. The tone profiles predicted from the stimulus discriminability values Δ_T plotted in Figure 2.3 under the assumptions that (1) each of the tones G, A, C, D, A♭ and D♭ produces activation that does not vary between major vs. minor contexts, (2) the major vs. minor thirds, sixths and sevenths produce opposite activation in major vs. minor contexts, and (3) the activations produced by G and D relative to the other tones match those in Figure 2.6.

Is performance in the semitone tasks mediated by a single mechanism?

In the previous section we used the term “neural apparatus” to refer to whatever auditory system component(s) listeners are using to perform the semitone tasks. This term was chosen to sidestep the question of whether or not this system is a single mechanism. As we have noted above, the separable model of Eq. (3) does a good job of accounting for most of the variance in the data. The separable model is also consistent with the idea that performance in the semitone tasks is mediated by a single mechanism of sort described by Eq. (2).

We suspect that this is not the case. Suppose Eq. (2) holds for some mechanism *M* with sensitivity function *f*. Then because $\Delta_2 = \Delta_3$ (as we see from Figure 2.3), it follows that either

$$f(\mathrm{B}) - f(\mathrm{B}\flat) = f(\mathrm{A}) - f(\mathrm{A}\flat) \tag{14}$$

or else

$$f(\mathrm{B}\flat) - f(B) = f(\mathrm{A}) - f(\mathrm{A}\flat). \tag{15}$$

If Eq. (14) holds, then a tone-scramble that contains 4 B's and 4 A♭'s (along with the usual 8 G_5's, 8 G_6's, 8 D's) will produce the same activation in *M* as a tone-scramble that contains 4 B♭'s and 4 A's. In this case, performance in a tone-scramble classification task in which Type 1 stimuli contain 4 B's and 4 A♭'s and Type 2 stimuli contain 4 B♭'s and 4 A's should be at chance for all listeners. On the other hand, performance should be equivalent to performance in the 2 and 3 tasks in a task in Type 1 tone-scrambles comprise 4 B♭'s and 4

A♭'s, and Type 2 tone-scrambles comprise 4 B's and 4 A's. If Eq. (15) holds, then performance should be at chance in the latter task, but in the former task performance should be equivalent to performance in the 2 and 3 tasks. Pilot data from two listeners indicates that performance is well above chance in both tasks. This suggests that listeners are making use of more than a single mechanism in performing the different semitone tasks. The sensitivity of all of these different mechanisms, however, seems to vary in parallel across different listeners.

Summary

139 listeners were tested in 6 different tone-scramble classification tasks. In each task, the listener strove (with trial-by-trial feedback) to classify the stimulus on each trial as Type 1 vs. Type 2 depending on the ensemble of tones it contained. In one of these tasks (the pitch height task), the stimuli differed much more strongly from each other than in the other 5 tasks. In each of the latter 5 tasks, called "semitone tasks," all tone-scrambles contained 8 G_5's, 8 G_6's, 8 D's to establish G as the tonic and D as the dominant of all tone-scrambles. In addition, Type 1 tone-scrambles contained 8 occurrences of one other tone *t* and Type 2 tone-scrambles contained 8 occurrences of the tone a semitone higher than *t*. In the 2, 3, 4, 6 and 7 tasks, the tone *t* was (respectively) A♭, B♭, C, E♭, and F.

Performance was much better in the pitch height task than in any of the semitone tasks. Performance (across listeners) was strongly correlated for all pairs of the semitone tasks; by contrast, performance in the pitch height task was correlated much less strongly with performance in the semitone tasks. This suggests that the pitch height task affords response strategies that are not useful for any of the semitone tasks.

The data were well described by a model in which the value of d' achieved by a given listener k in a given semitone task T was equal to $S_k\Delta_3$, where S_k reflects the listener's general sensitivity across the class of all semitone tasks and Δ_T reflects the discriminability of the Type 1 and Type 2 tone-scrambles used in semitone task T. Although this separable model did a good job of describing the structure in the data, a likelihood ratio test emphatically rejected the null hypothesis that this model reflects the true state of the world. However, the residual variance unaccounted for by the separable model did not show any clear structure. The Δ_T's for the five semitone tasks (see Figure 2.3) conformed to the following pattern: $\Delta_2 = \Delta_3 = \Delta_6 > \Delta_4 > \Delta_7$. The histogram of sensitivities S_k was strongly skewed in the positive direction (left panel Figure 2.4) leading to a bimodal distribution in the performance predicted by the separable model (right panel, Figure 2.4).

The 3 task had previously been shown to partition listeners into distinct high- and low-performing groups (Chubb et al., 2013). The primary motivation of the current study was to gain insight into the nature of the bimodally distributed sensitivity underlying performance in the 3 task. Is this sensitivity selective for differences between the major vs. minor modes, or is it more general? The current results support the claim that this sensitivity generalizes to a broad range of tone-scramble classification tasks requiring the extraction of qualities that depend on variations in the ensemble of tones present in a context polarized by a strongly established tonic. Only the 3, 6 and possibly the 7 tasks used in the current study might plausibly depend on sensitivity selective for differences between the major vs. minor modes. However, performance from these three semitone tasks was strongly correlated with performance in the 2 and 4 tasks, neither of which seems to require sensitivity to differences between the major vs. minor musical modes. Moreover, performance in the 2 task was equally as good as performance in both the 3 and the 6 tasks.

As argued in the discussion, it seems likely that the sensitivity that mediates performance in the semitone tasks is closely related to the sensitivity used by the listeners in the study of Krumhansl & Kessler (1982) in making the judgments that were used to derive the major and minor tone profiles. Finally, we argued that this sensitivity is unlikely to depend on a single mechanism; more plausibly, multiple mechanisms are used in different semitone tasks. If so, however, the current results suggest that the sensitivities with which these different mechanisms operate in a given listener are gated in parallel by a single process.

Chapter 3

Multidimensional Scaling of Mechanisms in Tonality

Abstract

This study investigated the ability for participants to pick out specific pitch classes among a large set of tonics and dominants. The stimuli once again consisted of tone scrambles that contained tones drawn from the equal tempered chromatic scale. Each scramble contained 18 tones: 3 G_4's, 3 D_5's, 3 G_5's, 3 D_6's, and 2 G_6's, as well as a single target tone. In any condition, each of the two participants listened for one of two target tones and was meant to respond with which of the two they believe they had heard. There were 55 conditions, one for every pair of target tones from the set of all pitch classes except for D. Sensitivities in the form of d' was calculated for each participant and condition and were arranged as a dissimilarity matrix for multidimensional scaling to put each tone into an n-dimensional space. 3 Dimensions were found for both participants. After rotating, scaling, and translating (operations that all preserve the relative distances among the tones), 3 seemingly meaningful dimensions were found for 1 participant and 2 were found for the other. Both of the participants appeared to have one dimension that arranged the tones from minor to major and one that arranged the tones from consonant to dissonant. Participant 2's third dimension arranged the tones by pitch height, and participant 1's third dimension was uninterpretable. This result is consistent with a past experiment by

Krumhansl (1979), in that it found the same number of dimensions, but dissimilar in that the tones were arranged differently.

Introduction

Music can elicit a range of emotionally colored perceptual qualities. While much work has been done on how tones fit together aesthetically, much about tonality is still unknown. As in chapter 2, we assume in this chapter that musical stimuli differ in quality only if they can produce differing levels of activation in one or more auditory perceptual "mechanisms." Much like the current study, several previous studies have put tones into a perceptual space (Krumhansl 1979, Krumhansl and Kessler 1982, Shepard 1982, Lerdahl 1988, Chew 2014). We once again imagine that the number of dimensions of the space is equal to the number of mechanisms that the brain has for tonality perception. For any two musical stimuli, A and B, we assume that the distance separating the centroids of activation produced in this space by A and B should correspond to the perceptual difference between A and B.

Particularly relevant to the current study is a study conducted by Krumhansl (1979). This study asked listeners to rate the similarity of different tone sequences to sets of two tones presented at the end of the sequences. From these ratings, she was able to find an average similarity rating of each combination of tones as it corresponded to the key of C major. She fit her data using multidimensional scaling and found it fit well to 3 dimensions. These tones formed a conical shape with tones existing in three parallel rings. In the smallest ring sat the tones that made up the C major chord, C, E, G, and the C in a higher octave. In the middle ring sat the tones in the C major scale that weren't in the C major chord, D, F, A, and

B. In the largest ring were the remaining tones, those not found in the C major scale, D♭, E♭, G♭, A♭, and B♭.

In the previous chapter, we looked at a large number of listeners and had them make several judgments involving tonal discrimination. While this was useful, the information we received was necessarily incomplete. We only looked at the perceptual differences between 6 pairs of note-ensembles. In the current study, we turn our focus to two musically experienced listeners in an attempt to take a more complete look at tonal discrimination. While in chapter 2 we look at a large number of listeners and only have them do a few tasks, in the current study we look at only two listeners, but have them discriminate more combinations of tones.

Methods

Stimuli

Each tone in the tone-scramble lasted 65 milliseconds. Each tone-scramble was 15 tones long for a total of .975 seconds. All tones were pure tones and drawn from the equal tempered scale and windowed using the following raised cosine function:

$$W(t)=\begin{cases}\dfrac{1-\cos\left(\dfrac{\pi t}{22.5ms}\right)}{2} & 0ms \le t < 22.5ms \\ 1 & 22.5ms \le t \le 42.5ms \\ \dfrac{1-\cos\left(\dfrac{\pi(t-65ms)}{22.5ms}\right)}{2} & 42.5ms < t \le 65ms\end{cases} \qquad (1)$$

In each condition, every tone-scramble included 15 tones, 14 of which comprised 3 G_4's, 3 D_5's, 3 G_5's, 3 D_6's, and 2 G_6's. In addition, each stimulus included one of two possible target tones T_1 or T_2. In "two-target" conditions, both T_1 and T_2 were drawn from the set, A, B, C, E, F, D♭, E♭, G♭, A♭, and B♭ (i.e., the set of all tones in the chromatic scale other than G and D). As usual, the order in which the tones were presented in a given tone-scramble was random. In every tone-scramble used in a two-target condition, the target tone was drawn randomly from the octave between G_4 and G_5 or between G_5 and G_6. In "one-target" conditions, T_2 was one of A, B, C, E, F, D♭, E♭, G♭, A♭, and B♭ drawn from the octave between G_4 and G_5 or between G_5 and G_6; however, T_1 was a G randomly chosen from among G_4, G_5, and G_6. In any given condition, on each trial, the listener strove to identify which one of T_1 or T_2 was included in the stimulus. Each listener was tested in 55 conditions (45 two-target conditions and 10 one-target conditions), one for every combination of two of the eleven target tones.

Participants

2 listeners participated. Both were UC Irvine students with self-reported normal hearing. The experimental procedures were approved by the UC Irvine Institutional Review Board.

Each listener was tested in each of the 55 conditions, one for each pair of the 11 notes. The listener was told that he/she would be attempting to identify which of the two target tones was presented to them.

In each condition, the listener was tested in blocks of trials, each containing 50 trials. Feedback was given visually after each trial; specifically, if the listener had responded correctly (incorrectly), the word "CORRECT" ("INCORRECT") appeared on the screen. After each block the message "% correct = X" appeared on the screen, where X was the listener's percentage correct in the most recent block. Listeners were verbally encouraged to take breaks in between blocks to avoid fatigue and keep performance levels as high as possible. Blocks and conditions were performed over the course of several weeks to minimize fatigue. For a given condition, the listener was trained until the listener felt that his or her performance had leveled out; then he or she ran 6 more blocks, and the data from those 6 blocks, comprising 300 trials total, were retained for analysis.

Results

We write $d'(T_1, T_2)$ for the value of d' achieved by a given listener in the condition with target tones T_1 and T_2. We modeled the data from a given listener by assuming that the listener responded "T_2" on a given trial j if

$$d'(T_1, T_2) + X_j > Crit(T_1, T_2) \tag{1}$$

where (1) $Crit(T_1, T_2)$ is the criterion used by the listener in the condition with target tones T_1 and T_2 and (2) the X_j's across trials j are jointly independent, standard normal random variables.

We take the $d'(T_1, T_2)$ values for both listeners across all conditions and put them into two dissimilarity matrices, one for each listener. These matrices are displayed in Table 3.1 with rows (columns) corresponding to target tones T_1 (T_2).

Listener 1											
	G	A♭	A	B♭	B	C	D♭	E♭	E	F	G♭
G		0.77	1.55	1.01	0.70	1.60	1.89	0.84	2.05	2.64	2.10
A♭	0.77		0.81	1.89	1.82	1.40	1.62	1.03	2.30	2.57	1.26
A	1.55	0.81		0.85	3.28	1.42	2.76	1.61	1.41	1.74	2.42
B♭	1.01	1.89	0.85		2.28	1.64	2.44	1.23	2.89	2.53	1.92
B	0.70	1.82	3.28	2.28		1.28	3.14	2.02	2.48	3.29	2.83
C	1.60	1.40	1.42	1.64	1.28		2.27	2.30	2.25	2.87	2.25
D♭	1.89	1.62	2.76	2.44	3.14	2.27		0.69	3.20	2.26	2.44
E♭	0.84	1.03	1.61	1.23	2.02	2.30	0.69		2.36	2.86	2.36

E	2.05	2.30	1.41	2.89	2.48	2.25	3.20	2.36		1.66	2.75
F	2.64	2.57	1.74	2.53	3.29	2.87	2.26	2.86	1.66		1.74
G♭	2.10	1.26	2.42	1.92	2.83	2.25	2.44	2.36	2.75	1.74	
Listener 2											
G		1.20	0.58	1.49	1.15	2.13	2.07	1.91	2.64	2.52	2.33
A♭	1.20		1.44	0.50	2.20	1.74	0.09	0.53	2.81	1.45	2.44
A	0.58	1.44		0.74	0.81	0.47	2.02	2.53	2.42	2.04	2.40
B♭	1.49	0.50	0.74		3.42	1.82	1.73	0.68	3.24	3.02	2.07
B	1.15	2.20	0.81	3.42		0.77	2.16	2.89	0.12	3.36	0.69
C	2.13	1.74	0.47	1.82	0.77		1.57	0.60	0.98	1.41	1.54
D♭	2.07	0.09	2.02	1.73	2.16	1.57		0.30	2.20	2.66	1.96
E♭	1.91	0.53	2.53	0.68	2.89	0.60	0.30		2.86	1.98	2.41
E	2.64	2.81	2.42	3.24	0.12	0.98	2.20	2.86		2.50	0.34
F	2.52	1.45	2.04	3.02	3.36	1.41	2.66	1.98	2.50		0.79
G♭	2.33	2.44	2.40	2.07	0.69	1.54	1.96	2.41	0.34	0.79	

Table 3.1. The obtained $d'(T_1, T_2)$ values for listeners 1 and 2. All Notes of the chromatic scale were paired against all other notes in the chromatic scale, with the exception of D. Listeners performed very poorly in some tasks, while doing exceptionally well in others.

We ran the Matlab nonparametric multidimensional scaling function, mdscale, on each of the matrices, to derive an arrangement of points in all numbers of dimensions from one to six (seven dimensions yielded no better fit than six). Stress levels for each participant can be found in Table 3.2.

# of Dimensions	1	2	3	4	5	6
Listener 1	0.3722	0.1789	0.0947	0.0494	0.0333	0.0069
Listener 2	0.3451	0.1373	0.0883	0.0442	0.0201	0.0023

Table 3.2. The obtained stress levels for each dimensional fit for each listener. Stress values for both participants are very similar for all dimensions. We used the three-dimensional fit for all subsequent analyses.

We picked the three-dimensional fit for several reasons. Stress values for listeners 1 and 2 for these solutions were 0.0883 and 0.0947 respectively. This is very close to the fit used in the Krumhansl (1979) experiment, which was 0.108. This was also the same number of dimensions found in the Krumhansl experiment.

We transform each listener's data using translations, reflections, and orthogonal rotations (which will preserve the relative distances among all of the points), such that the data sit on axes that appear meaningful. Table 3.2 contains the solutions to the three dimensional fits. Figure 3.1 shows a three-dimensional plot of listener 1's three-dimensional solution, and Figure 3.2 shows a three-dimensional plot of listener 2's three-dimensional solution.

	Listener 1				Listener 2		
	Dim 1	Dim 2	Dim 3		Dim 1	Dim 2	Dim 3
G	-0.07	-0.83	-0.41		0.08	-0.34	-1.74
A♭	-0.14	0.14	-0.67		-1.19	0.37	-0.21
A	-0.07	-0.05	0.92		0.27	0.62	-1.30
B♭	-0.26	-1.07	0.63		-1.24	-0.51	-0.88
B	1.17	-0.72	-1.14		1.61	-0.41	-0.47
C	0.95	-0.76	-0.02		0.15	0.26	0.40
D♭	-1.49	0.15	-0.69		-0.91	-1.18	0.53
E♭	-1.10	-0.37	-0.02		-1.56	-0.35	0.83
E	0.94	0.80	1.16		1.68	-0.75	0.84
F	-0.21	1.34	1.02		-0.14	1.86	0.88
G♭	0.28	1.37	-0.78		1.24	0.43	1.13

Table 3.3. The obtained points from multidimensional scaling the $d'(T_1, T_2)$ values using 3 dimensions and then rotated, translated, and reflected for listeners 1 and 2.

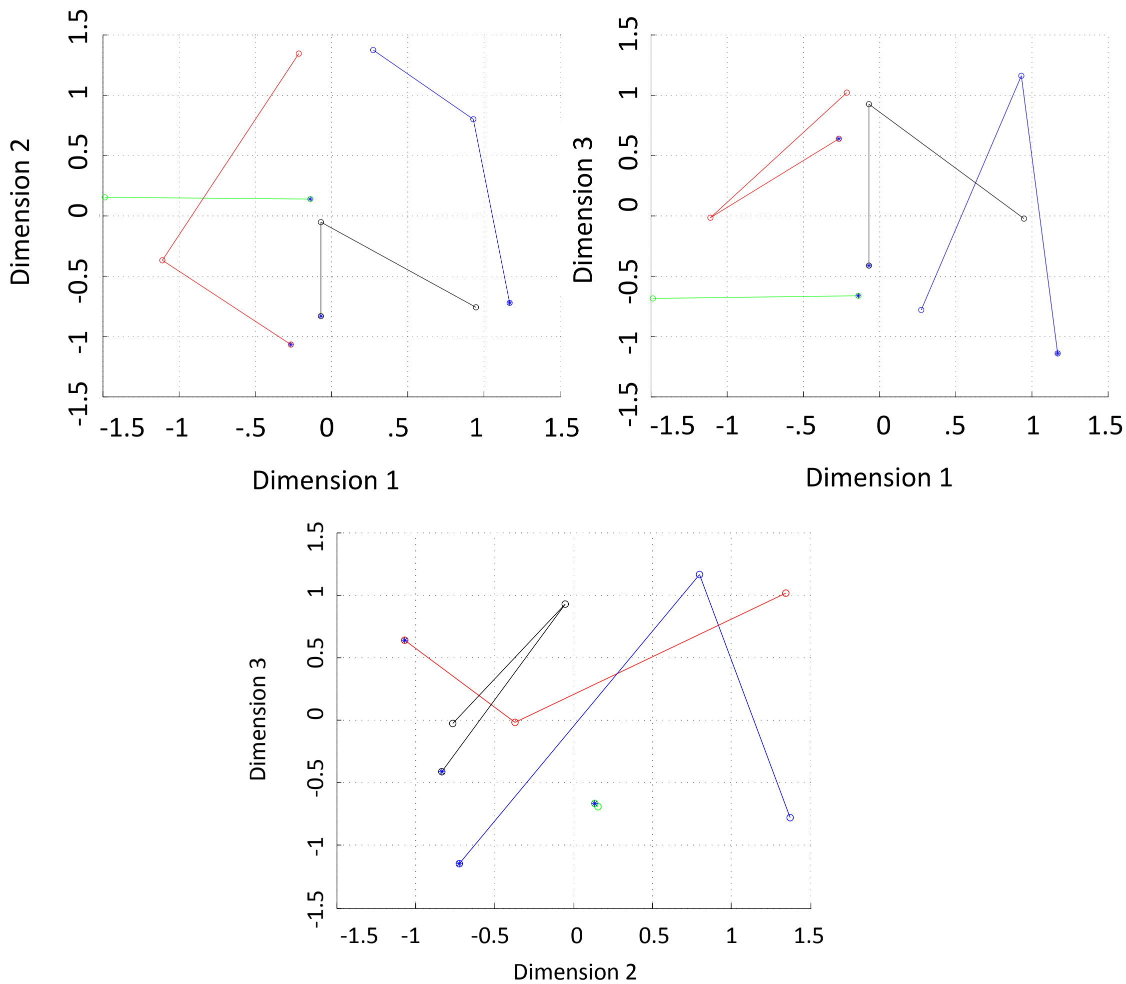

Figure 3.1. The multidimensional scaling fit and rotation in 3 dimensions for listener 1. The black lines represent the tones in both G major and minor modes: G, A, and C, respectively. The blue lines represent the notes in the G major scale only: B, E, and G♭, respectively. The red lines represent the notes in the G (natural) minor scale only: B♭, E♭, and G, respectively. The green lines represent the notes found neither in the G major nor G minor scale: A♭ and D♭, respectively. All sets of notes begin with the node marked with a blue star and are ordered in their sets from lowest to highest. Dimension 1 can be found on the abscissa of both of the upper graphs. Dimension 2 can be found on the ordinate of the

upper left graph and the abscissa of the lower graph. Dimension 3 can be found on the ordinate of the upper right graph and the lower graph.

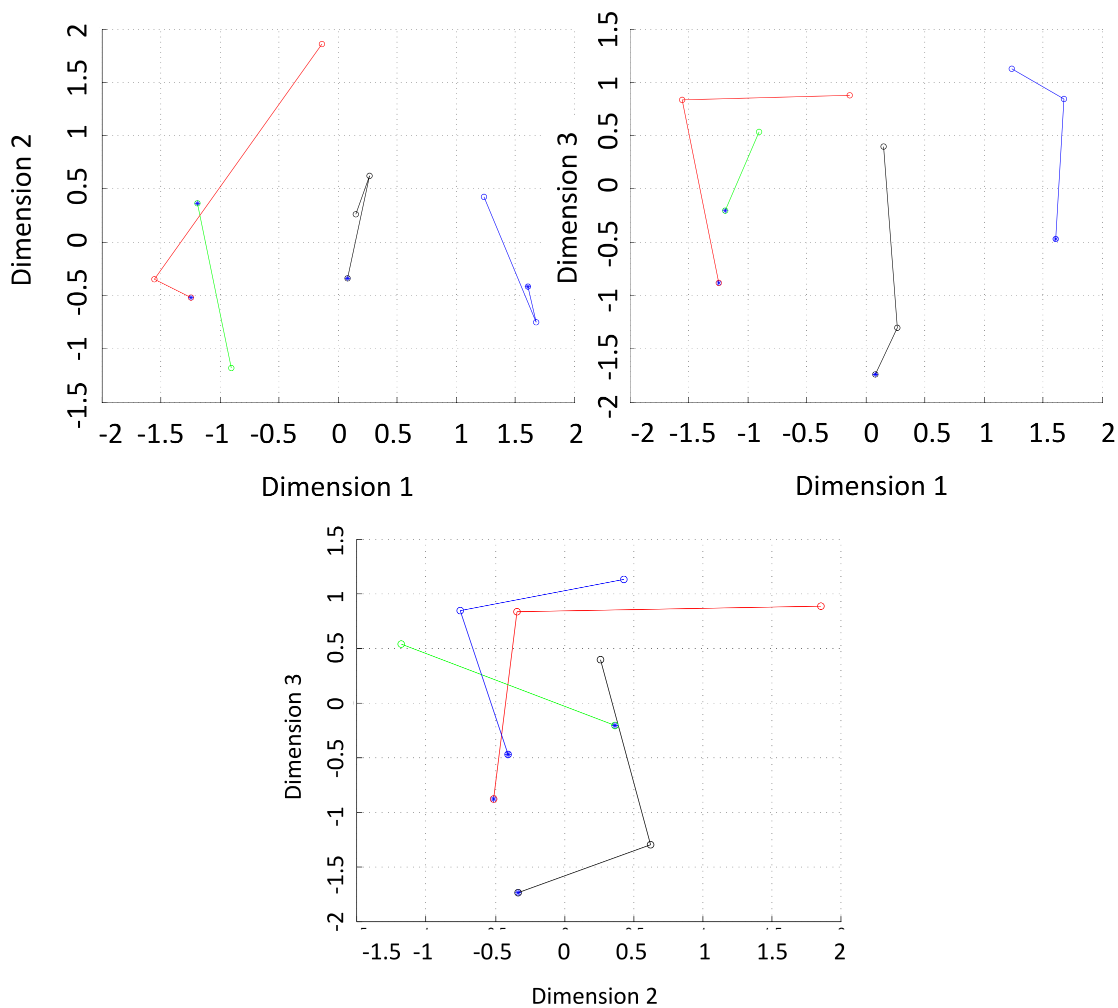

Figure 3.2. The multidimensional scaling fit and rotation in 3 dimensions for listener 2. The key is the same as Figure 3.1.

The multidimensional fits for listener 2 show that the tones group into 3 clusters. In one cluster are the tones that exist in both the major and minor scales, G, A, and C (shown in black). In another cluster are the tones that only exist in the major scale, B, E, and G♭

(shown in blue). In the third cluster are both the notes that show up in only the minor scale, B♭, E♭, and F (shown in red), and the notes that do not exist in either scale, A♭ and D♭ (shown in green).

Dimension 1 for listener 2 does a good job distinguishing the major notes from the minor notes. Though this is the case, a few things come out of using this dimension as differentiating major and minor. While the notes found only in the major scale (B, E, and G♭) tend to clump toward the high end of this dimension, and notes found in both scales seem to clump in the middle, the minor notes do not distinguish themselves quite as clearly. Note a minor seventh above the tonic, F, seems to be closer to the neutral notes on this axis. In addition, the notes found in neither scale tend to be much closer to two of the notes that are only found in the minor scale, and are much lower one dimension 1 than the other tones.

Listener 1 shows a similar pattern for dimension 1. While listener 1's first dimension does a pretty good job distinguishing the major and minor notes, the tones certainly do not clump quite as clearly as they do for listener 2. The notes found only in the major scale (B, E, and G♭) are indeed found on the far higher end of dimension 1, but the tone a perfect fourth above the tonic also is as high as the major tones are. The remaining tones found in both scales (G and A) are the next largest on this dimension. The remaining tones (A♭, B♭, D♭, E♭, and F) are all the lowest on this dimension. While these notes do not fall in the same order as they do for listener 2, nor do they clump quite as clearly, the groupings found in listener 1 still exist mostly in the same order, with scale degree four, C, being the one exception.

The second dimension for listener 2 seems to correspond mostly with the notion of consonance and dissonance. Toward the lower end of dimension 1, we see G, B♭, B, E, and E♭. These are all tones that form a consonant interval with scale degree the tonic, G. Higher up on the graph are A♭, A, C, F, and G♭. All of these tones, with the exception of C, form a dissonant interval with the tonic. C is a perfect 4th above the tonic, which is sometimes considered a dissonant interval and sometimes a consonant interval. Interestingly enough, it takes the value closest to the consonant interval of the dissonant tones. The one exception to this in our fit is D♭. D♭ forms a tritone with the tonic, which is a dissonant interval, and this tone is found at the consonant end of our consonance-dissonance axis.

Listener 1 shows a similar consonant-dissonant axis for their dimension 2. G, B♭, C, E♭, and E all form consonant intervals with the tonic, G. Once again, C can be considered a consonant or a dissonant interval. The notes higher on this axis, A♭, A, D♭, F, and G♭, form a dissonant interval with the tonic. The one exception in this case is the tone B, which is consonant, but sits with the dissonant tones.

The third axis for listener 2 seems to order the notes by pitch height. The notes in this dimension from bottom to top are: G, A, B♭, B, A♭, C, D♭, E♭, E, F, G♭. These notes are entirely in pitch height order, except for A♭.

The third axis for listener 1 does not seem to show any kind of pattern.

Discussion

With a few exceptions, the fits for the multidimensional scaling fits show 3 dimensions that correspond to major-minor, consonance dissonance, and for one listener, pitch height. While these fits are similar to Krumhansl's (1979) fits in that they contain three dimensions, and about the same levels of stress, they do not conform to quite the same shape as her study's points did. In her study, the points formed three circles each larger than the other, as though wrapped around a cone. The tones from the major chord were along the smallest circle, the remaining tones in the major scale were on the middle circle, and the non-scale tones were placed along the largest circle. We did not find that shape.

However, our study differs from hers in several key ways. Most importantly, our study did not set the mode to major as hers did. While we did establish a tonic, we did not establish a mode (that could differentiate major and minor) in our study. It is very reasonable to think that the tones' relationship to one another in the context of a specifically major key as opposed to their relationship in a modally ambiguous key could be very different. In this case, that seems to be true. In particular, she finds a hierarchy in her fit, and that is very dependent on mode. The arrangement she finds definitely seems that it would not hold for a minor key, as the notes she finds closest to the tonic are in the major key, and we would likely find the opposite if a minor context was used.

The second critical difference between our study and Krumhansl's is the measure for note closeness. While we use sensitivity to note differences, she used a rating scale of similarity. This may also affect the judgments in different ways. Much of her data may be the result of listeners' a priori beliefs about the way music should sound or what they have learned about music. However, because our judgments are the result of what we assume to be stochastic processes, we suspect there may be a fair amount of noise in our data. This may explain why the dimensions don't quite form perfect patterns where we see them.

Summary

This study used tone-scrambles (rapid, random sequences of tones) to investigate the mechanisms conferring sensitivity to variations in the ensemble of notes comprising the scale relative to a fixed tonic. On each trial in every condition, to establish G as the tonic, the stimulus on each trial was a tone-scramble comprising 3 G_4's, 3 D_5's, 3 G_5's, 3 D_6's, and 2 G_6's. In addition, in a given condition, there were two target tones, T_1 or T_2, and the tone-scramble presented on any trial contained a single tone that was either T_1 or T_2. The task was to judge (with feedback) which of T_1 or T_2 had been present in the stimulus.

Each of 2 listeners was tested in each of 55 different experimental conditions. In 10 "one-target" conditions T_1 was a G (randomly chosen from among G_4, G_5, and G_6), and T_2 was one of A, B, C, E, F, D♭, E♭, G♭, A♭, and B♭ (randomly chosen on each trial from the octave between G_4 and G_5 or between G_5 and G_6). In 45 "two-target" conditions, each of T_1 and T_2 was one of A, B, C, E, F, D♭, E♭, G♭, A♭, and B♭. 300 trials were obtained from each listener in each of these 55 conditions. Prior to data collection in a given condition, the listener received sufficient training in that condition to reach asymptotic performance.

Performance in the condition using a given pair of target tones, T_1 and T_2 was summarized by a value $d'(T_1, T_2)$ reflecting the sensitivity of the listener to the difference between the two sorts of tone-scrambles used in that condition, and hence between the tones T_1 and T_2 in the context of strongly established tonic of G. Multidimensional scaling was used to assign the tones G, A, B, C, E, F, D♭, E♭, G♭, A♭, and B♭ (all the tones of the chromatic scale relative to the tonic G other than the dominant D which was not varied in this experiment) to points in spaces of various dimensions so that the Euclidean distances between points corresponding to pairs of notes (T_1, T_2) approximated $d'(T_1, T_2)$ with minimal stress.

The lowest dimension providing a reasonable fit for each of the two listeners was 3. Under a particular rotation of the MDS solution, one "major-minor" dimension separated tones of the major vs. the minor scale. A second "consonance-dissonance" dimension separated notes that form a consonant interval with the tonic G from tones that form dissonant intervals with G. The third dimension was uninterpretable for one listener but sorted tones in order of pitch-height for the other listener.

Chapter 4

Interactions Between the Tonic and Major and Minor Tones in Major vs. Minor Judgments

Abstract

This study investigated the basis of the emotional qualities that are produced by music with a tonic center. Specifically, this experiment was designed to investigate the interaction between the strength with which the tonic is established in a sequence of tones and the perceived major vs. minor affect of the sequence. The stimuli used were tone-scrambles: quickly presented, randomly ordered sequences of pure tones, drawn from the equal-tempered chromatic scale. All tone-scrambles used contained 18 tones, and used the tonic G that was established strongly throughout the experiment. On any given trial, the participant strove to judge whether the stimulus contained targets that were B's and E's (making it major) vs. B♭'s and E♭'s (making it minor). Scrambles were made up of varying numbers of tonics (between zero and seven), varying numbers of targets (between zero and seven), and 3 copies of the dominant D; the remaining tones (to make the total number 18) were other, randomly chosen tones from the chromatic scale. Participants were good at the task, but showed substantial influence of tones they were not meant to listen to. In particular, both listeners were influenced differently by the last tone of the sequence than they were by the other tones. Participants also showed an asymmetric interaction between tonic strength and major vs. minor target tones: the influence of both major and minor

target tones was enhanced in tone-scrambles with high tonic strength; however, this effect was more pronounced for minor than for major target tones.

Introduction

The major and minor diatonic scales are central to the Western music. Music using the major scale is considered to sound happy, and music using the minor scale is considered to sound sad. Many past studies have shown that participants do indeed feel this way (Crowder, 1984, 1985a,b, Gagnon and Peretz, 2003, Gerardi and Gerken, 1995, Heinlein, 1928, Hevner, 1935, Kastner and Crowder, 1990).

The qualitative differences between music using the major vs. minor scale depend critically on the establishment of a “tonic,” i.e., a fundamental note of the scale that serves to center the music in a specific key. Thus, for example, the prevailing quality extracted by a listener from a musical sequence comprising various proportions of only the notes C, D, E, F, G, A, and B depends on which (if any) of these seven notes is cast in the role of the tonic. If the note C occurs more frequently and with greater volume and rhythmic emphasis than other notes, then C is likely to be heard as the tonic, and the music will take on the “happy” quality characteristic of the C major scale. On the other hand, if A is emphasized instead of C so that A comes to be heard as the tonic, then the music will take on the “sad” quality of the A (natural) minor scale. If a different one of these seven notes is emphasized as the tonic, then the prevailing quality of the music will differ from each of these in a way characteristic of one of the modes. For example, if F is emphasized, then the quality produced will be characteristic of the Lydian mode.

The tonic is also important because it typically operates in various ways to structure the sequence of a melody. For example, melodies typically end on the tonic because this imbues the melody with a sense of finality that other tones fail to provide.

The present study focuses not on this aspect of musical experience, however, but rather on how the presence of a tonic operates to enable other notes of the scale to inject mood-like coloration into the qualities evoked by the music. Though composers intuit well how to produce those qualities, the source from which these qualities spring is still unknown.

Music that relies on a tonic center is essentially *interactional*. The tonic interacts with the other notes of the scale to produce the qualities evoked by the music. The impact exerted on the quality of an unfolding piece of music by a B depends crucially on which note has been established as the musical tonic. If the tonic is G, then B is the major third of the scale, and it is likely to inject a burst of "happiness" into the music. By contrast, if the tonic is G#, then B is the minor third of the scale, and it is likely to inject a jolt of "sadness."

It is the nature of this interaction that is the focus of the current chapter. Of central interest is the question of whether the notes of the minor scale interact with the tonic in a different way than do the notes of the major scale. Any such difference seems likely to shed light on the mysterious emotional qualities produced by these two different modes. To investigate this question, we use a task in which the listener is required to classify (with feedback) brief, proto-musical stimuli according to whether they the contain major thirds and major sixths vs. minor thirds and minor sixths (the particular scale components that operate most strongly to produce the emotional qualities distinctive of the major vs. minor modes). In addition, in several different intermixed conditions, we introduce strong trial-by-trial variations in the number of tonics present in different stimuli.

Methods

Stimuli

Each participant was tested in 50 Blocks of 48 trials. Stimuli consisted of 32 randomly ordered tones, each of which was 65ms in duration with the entire tone-scramble lasting 2.08 seconds. All tones were pure tones drawn from the equal tempered scale and windowed using a raised cosine function:

$$W(t) = \begin{cases} \dfrac{1-\cos\left(\dfrac{\pi t}{22.5ms}\right)}{2} & 0 \le t < 22.5ms \\ 1 & k \le t \le 42.5ms \\ \dfrac{1-\cos\left(\dfrac{\pi (t-65ms)}{22.5ms}\right)}{2} & 42.5 < t \le 65ms \end{cases} \quad (1)$$

The stimuli in any given block were tone-scrambles generated by 8, randomly interleaved staircases, S_k, k=0,1,...,7. Each tone-scramble generated by staircase S_k contained exactly k tonics; the number of these that were G_5's was equally likely to be 0, 1, 2,..., k, the rest being G_6's. In addition, each tone-scramble (regardless of which staircase generated it) contained three D_6's minimum, with more potentially added as described below. The "signal level" $L_k(t)$ of the stimulus generated on the t^{th} trial presented by staircase S_k was 7 for t = 1 and was controlled thereafter by a 3-down, 1-up rule. Specifically, if the listener had responded correctly on trials t-3, t-2, and t-1 presented by S_k, then $L_k(t) = L_k(t - 1) - 1$ (provided $L_k(t - 1) > 1$); otherwise, $L_k(t) = L_k(t - 1) + 1$ (provided $L_k(t - 1) < 7$). In any given block, there were 24 major and 24 minor trials, presented in a random order, such that on any given trial, the tone-scramble presented was equally likely to be major or minor. If the stimulus was major, then (1) it contained a number of E_6's that was equally likely to be any of 0, 1,..., $L_k(t)$, a number of B_5's equal $L_k(t)$ minus the number of E_6's; the rest of the stimulus tones (of which there were 18-k-3- $L_k(t)$) were drawn randomly, with replacement, from the set {$A\flat_5$, A_5, C_6, $D\flat_6$, D_6, F_6, $G\flat_6$}. If the stimulus was minor, then it contained $E\flat_6$'s instead of E_6's and $B\flat_5$'s instead of B_5's. Note that a major stimulus was devoid of minor thirds ($B\flat_5$'s) and minor sixths ($E\flat_6$'s), and a minor stimulus was devoid of major thirds (B_5's) and major sixths (E_6's).

Participants

2 listeners participated. Both were well trained in several tasks similar to this, involving major and minor judgments using tone-scrambles. Both participants had self-reported normal hearing. The experimental procedures were approved by the UC Irvine Institutional Review Board.

The Experiment

Participants were instructed to listen to each tone scramble and judge whether it contained major or minor targets. The participant was tested in fifty blocks of trials in this same condition, each containing 48 trials (for a total of 2400 trials). The type of the stimulus on any given trial equally likely to be 1 or 2; each participant received an equal number of type 1 and type 2 trials in every block. Staircases began at the easiest level, with 7 targets in a single scramble. Feedback was given visually after each trial; specifically, if the participant had responded correctly (incorrectly), the word "CORRECT" ("INCORRECT") appeared on the screen. Participants were encouraged to take breaks in between blocks to avoid fatigue and keep performance levels as high as possible. The 50 blocks were completed over the course of several days, at the pacing of the participant.

Modeling

We write $N_j(t)$ for the note in the t^{th} pip of the j^{th} tone-scramble presented to the listener. We assume that the observer responds "major" on trial j (and "minor" otherwise) if

$$\Gamma_j + X_j > Crit \tag{2}$$

where *crit* is the decision criterion used by the listener, the X_j's used across trials *j* are jointly independent, standard normal random variables, and

$$\Gamma_j = \sum_{t=1}^{17} W(t)F\left(N_j(t)\right) + W(18)F_{last}\left(N_j(18)\right) + I_j \tag{3}$$

where

1. Each of F and F_{last} is a real-valued function of the 13 notes of the equal-tempered chromatic scale from G_5 to G_6 constrained to sum to 0 and to have a sum of squared values equal to 1;
2. W is a real-valued function of the set of pip locations {1, 2, …, 18} constrained to have nonnegative sum approximated by a third degree polynomial;
3. I_j is a term that depends on (1) the strength with which the tonic is established in tone-scramble *j* as well as (2) the signal in tone-scramble *j*. Specifically, the strength with which the tonic is established in tone-scramble *j* is gauged by

$$TonicStrength_j = \sum_{t=1}^{18} W(t)T_j(t) \quad \text{for} \quad T_j(t) = \begin{cases} 1 & \text{if } N_j(t) \text{ is a G} \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

And the signal is gauged by

$$Signal_j = \sum_{t=1}^{18} W(t) S_j(t) \tag{5}$$

for

$$S_j(t) = \begin{cases} F\left(N_j(t)\right) & \text{if } t < 18 \text{ \& } N_j(t) \text{ is not G} \\ F_{last}\left(N_j(t)\right) & \text{if } t = 18 \text{ \& } N_j(t) \text{ is not G} \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

Then we define I_j in terms of the model parameters $\rho_1, \rho_2, \ldots, \rho_8$ as follows. Let

$$\xi_j = \text{TonicStrength}_j \times \text{Signal}_j, \tag{7}$$

and let Ω_ξ be the set of values taken by ξ_j across all trials *j*. Then for $k = 1, 2, \ldots, 8$,

$$I_j = \rho_k \text{ if } \xi_j \text{ is in octile } k \text{ of } \Omega_\xi. \tag{8}$$

The ρ_k's are constrained to sum to 0 (otherwise, they can trade of with *Crit*).

In this model, we can think of any given pip as exerting an impact on the listener's response on trial *j* that is a separable function of the note of that pip, as well as the pip's location in time. Thus, for example, the primary influence exerted on the response of the participant on trial *j* by the note of any pip t < 18 is given by the term $W(t)F\left(N_j(t)\right)$ in the leftmost sum of Eq. 3. (The qualifier "primary" is necessary in the previous sentence because the tone at pip t can also influence the response of the participant indirectly by influencing the interaction term I_j in Eq. 3 above.) For any pip t, $W(t)$ reflects the sensitivity of the listener to tones that occur at pip t, and for any note n, $F(n)$ gives the relative influence exerted by tones with note n on the listener's judgments. When $W(t)F\left(N_j(t)\right)$ is positive (negative), it causes the listener to be more likely to say "major" ("minor"). The function F is constrained to sum to 0; otherwise it could trade off with the model parameter *Crit*. F is also constrained to have a sum of squared values equal to 1; otherwise it could trade off in amplitude with W. We expect $W(t)$ is to approximate a smooth function, as there should not be a large difference between a pip and the pip just after it. Therefore, we approximate $W(t)$ as a third degree polynomial.

The term $W(18)F_{last}\left(N_j(18)\right)$ in Eq. 3 gives the primary impact exerted by the last pip of the tone-scramble on trial j on the judgment of the listener. As in the case of any pip $t < 18$, $W(18)$ reflects the sensitivity with which the listener's responses are influenced by notes that occur on pip 18. We anticipate, however, that the last (18th) pip may exert a different pattern of influence on the listener's response than the other (non-final) pips in the tone-scramble. The final notes in melodies tend to obey different rules than the notes leading up to them. In particular, many melodies take the tonic as their last note. This is why we allow the function, F_{last} (that defines the relative influence exerted by the note of the last pip) to differ from the function F (that defines the relative influence exerted by the notes of non-final pips). As will soon be evident, F_{last} does in fact differ significantly in form from F.

Finally, we introduce an interaction term I_j that can be viewed as providing a correction to the first two terms in Eq. 3 above. We anticipate that the strength with which the tonic is established in a given tone-scramble is likely to modulate the influence exerted on the listener's judgment by the other (non-tonic) notes. In particular, we expect that higher tonic strength on a given trial will heighten sensitivity to B's and E's (the major target notes) as well as to B♭'s and E♭'s (the minor target notes). A key question is: will the heightened sensitivity conferred by increased tonic strength be symmetric for major vs. minor target notes? An asymmetry in the pattern of this modulation for major vs. minor target notes might provide a clue concerning the source of the strikingly different qualities evoked by major vs. minor music.

We take a nonparametric approach to modeling this interaction. We first define the *tonic strength* in a given tone-scramble by adding up the weights $W(t)$ for all pips t to which a tonic has been assigned (Eq. 4). We then define the *signal* in the tone-scramble by adding up the influences exerted by all pips to which any note other than a tonic has been assigned (Eqs. 5 and 6). Next, we define the *interaction strength* of a given tone-scramble by multiplying its *tonic strength* times its *signal* (Eq. 7). We then order the interaction-strength values of all tone-scrambles from lowest to highest and put the interaction values into eight equal bins. All tone-scrambles j whose interaction-strengths fall into bin k are assigned $I_j = \rho_k$ (Eq. 8).

The participant takes the sum of each of the 3 terms in Eq. 3, which we think of as reflecting the net major or minor impact of the entire tone-scramble. We then add some noise to the judgment, and that gets compared to some criterion that the listener uses, which we assume is constant across all trials. If the value is higher than the criterion, the participant responds “major,” and otherwise, they respond “minor.”

This model has 34 degrees of freedom. There are 13-2 degrees of freedom in the function F that gives the relative impact of each note when is assigned to any pip other than the last one. There are 13-2 degrees of freedom in the function F_{last} that gives the relative impact of each note when it is assigned to the last pip. There are 4 degrees of freedom in the function W (the coefficients of the four terms of the polynomial including the intercept), which gives the sensitivity of the listener to notes occurring at each of the 18 pips found in all of the tone-scrambles. There are 8-1 degrees of freedom in the interaction function, I_j, one parameter for each of the bins for the interaction values. There is one parameter for the criterion used by the listeners.

Fitting methods

For each participant, a Markov-chain Monte-Carlo simulation is used to derive a sample size of 50,000 from the posterior density characterizing the joint distribution of all the model parameters. Each parameter was given a prior uniform density on the interval (-10,10), which was several times larger than the 95% credible intervals found for any of the parameters.

Results

The fits to the model can be seen in Figures 1, 2, and 3 for both participants.

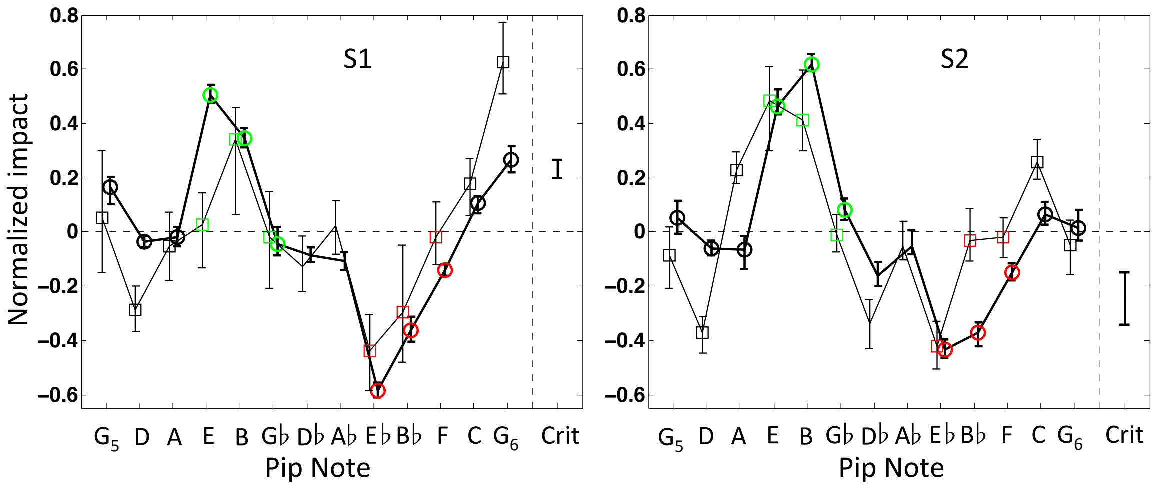

Figure 4.1. The tone functions for participants 1 and 2 on the left and right respectively. The darker line and circles display the mean and error bars for the parameters in the function ?, and the lighter line and squares display the parameters for the function F_{last}. The notes are ordered in the clockwise direction of the circle of fifths. The black circles and squares show tones that are part of both the major and minor scales. The green circles and squares show tones that are a part of the major scale only, and red circles and squares show tones that are a part of the minor scale only. Notes without circles are a part of neither scale.

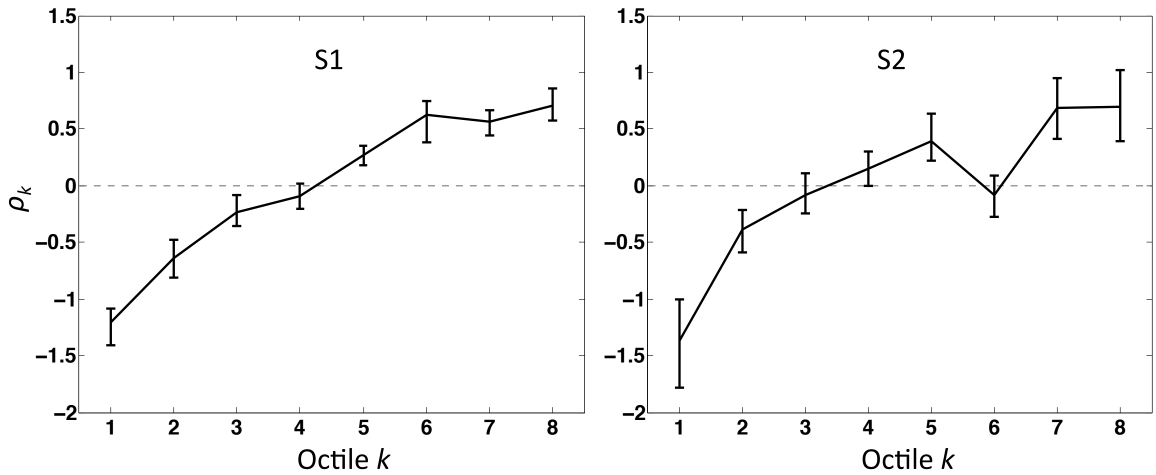

Figure 4.2. The interaction parameters ρ_k corresponding to interaction level octiles k = 1,2,...,8 for participants 1 and 2 on the left and right respectively.

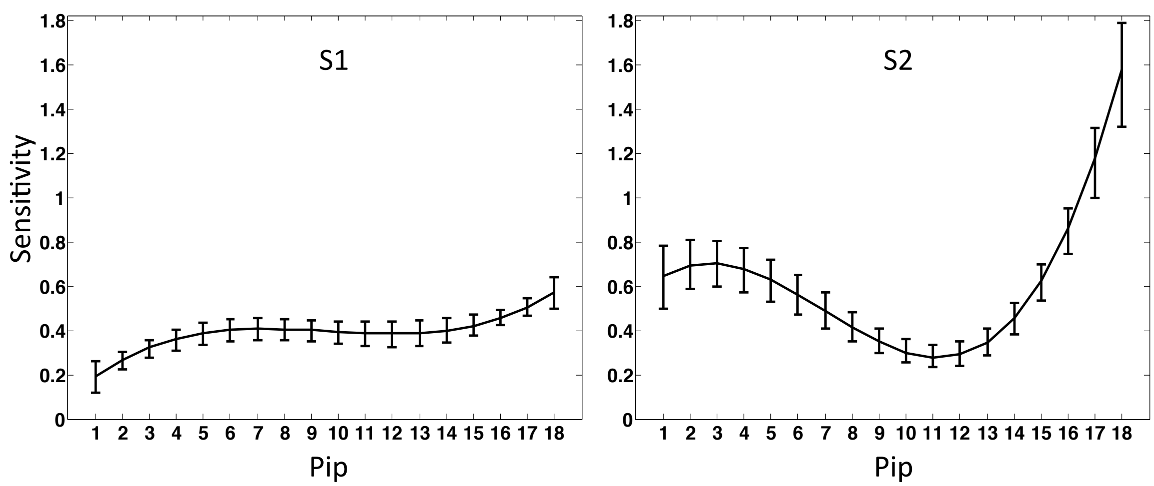

Figure 4.3. The temporal weights for participants 1 and 2 on the left and right respectively. The temporal weights are assumed to conform to a 3rd degree polynomial.

The tone weights for the non-final tones in Figure 4.1 show that the participants are quite good at this task, as they are giving the most positive weight to tones E and B, while giving the most negative weight to the tones E♭ and B♭. While the most weight was given to the target tones in this task, those tones were not the only tones that exerted a nonzero influence on the listeners. D, A, D♭, and F all exerted a minor influence for participant 1 while G♭, and C both exerted a major influence. D♭, A♭, and F all exerted a minor influence on participant 2 while G_5, C, and G_6 all exerted a major influence.

In addition, the tone weights for the final tones in Figure 4.1 show the participants indeed often have different weights for the last tone compared to the non-final tones. This is especially notable in tones D and C for both participants, which have near neutral influences when non-final, but exert minor and major influences respectively when they are the final tone in a sequence. The differences are also consistent between the participants in note F, which is minor when non-final, but neutral when heard at the end of a sequence.

The interaction parameters differ significantly from zero implying that the strength with which the tonic is established in a given tone-scramble does indeed modulate sensitivity to major and minor target tones in the stimulus. The predominant increasing, linear trend in each of the functions plotted in Figure 4.2 indicates that as tonic strength increases, sensitivity increases to both major influences as well as minor influences.

Moreover, this interaction effect appears to be asymmetrical with respect to major vs. minor influences. The very low values of interaction seem to be giving the participants more minor impression than the very high values of interaction, as ρ_1 for both participants is around -1.25 whereas ρ_8 for both participants is about .75.

For t = 1,2,…, 18, the values $W(t)$ are all above zero for both participants. This shows that the participants are indeed using the information they are hearing in the tone scrambles, and they are getting information from all points in the scramble. However, it does not seem like they are weighting all points in the scramble equally. In particular, it seems as though the participants are weighting the last pip in the sequence higher than the rest. This is especially true for participant 2.

Discussion

If we believe that the participants have a mechanism that allows them to listen for major or minor, this mechanism is not affected by only the major and minor notes. If it were the case that this mechanism keyed into only the major and minor tones, then we would expect the weight of the non-major and non-minor notes (G_5, D, A, D♭, A♭, C, G_6) to be zero. This is not the case. Other notes that are not traditionally thought of as major or minor do impact this judgment.

Also of interest are the notes a major 7th and minor 7th above the tonic (G♭ and F in our experiment). Our past experiments have shown that scale degrees 3 and 6 behave extremely similarly to one another. In the canon of western music, scale degrees three and six are both unambiguous indicators of whether the scale is major or minor; however, scale degree seven doesn't share this property. The harmonic minor scale contains a sharped scale degree seven, while the natural minor scale does not, so the presence of a note a major seventh above the tonic is not necessarily indicative of being in a major key. However, the presence in the scale of a minor seventh does unambiguously indicate that the scale is minor. With this information, we might think that F would have a minor weight, indicating that we are in a minor key, but that G♭ might have less major weight or even neutral weight. This is indeed the pattern that we see in the function *F* for both listeners in Figure 4.1.

Also of interest is the minor effect of the tones a minor second and a tritone away from the tonic (A♭ and D♭ in our experiment). Both of these notes cause the listeners to be more likely to respond "minor." This is somewhat strange, as neither tone is in either the major or minor scales. Because of this, one might think that these notes should impact the listener in neither direction. We may get this result because, though these notes exist in neither scale, they are very dissonant intervals. It is possible that the participants are not entirely able to disentangle the sensation of dissonance and minor. While these are certainly not the same concept, they both are associated with a negative affect, and it certainly doesn't seem unreasonable that the sensations between them are at least somewhat similar.

Somewhat oddly, we find that the last tone of the sequences impacts the listeners quite differently than the notes in the rest of the scrambles. This may be related to the notion of resolution, often discussed in music theory. While the fifth scale degree (D in this experiment) usually has an almost neutral impact, for both listeners it exerts a negative impact (i.e., an impact inclining the listener to respond "minor") if it occurs on the last pip of a tone-scramble. Conversely, scale degree four (C in this experiment) exerts a positive impact when it occurs on the last pip. In the same vein, for listener S2, the note a tritone above the tonic (D♭) exerted negative influence when it occurred on the last pip, and for listener S1, the high octave (G_6) exerted a positive influence when it occurred on the last pip. These observations suggest that if the scramble sounds especially unresolved at the end, this can exert a negative (toward minor) influence, and if the scramble sounds especially resolved, this exerts a positive (toward major) influence.

More evidence for the notion of resolution being important to major and minor judgments can be found in the function *F* reflecting the relative influence of non-final tones. Both participants show a slight major influence of scale degree 4 (C) and a slight minor influence of scale degree 5 (D). Tonal music is usually said to be moving toward a tonic. In addition, chords often move in the counterclockwise direction of the circle of fifths, or downward by fifths (backwards along the abscissa in Figure 4.1). This puts the chord based on scale degree 4 right after the tonic, closest to the beginning. Likewise, the chord based on scale degree 5 is right before the tonic and closest to the end. Sequences ending on scale degree 5 are usually thought of as being especially unresolved. It may be the case that when scale degree 5 appears in a non-final pip this injects a small amount of negative impact related to the unresolvedness it would create if it occurred on the last pip.

We also find an interaction between the major impact of a tone-scramble and the number of tonics that a tone-scramble contains. This is not entirely surprising, as the establishment of a key is very important for being able to identify the quality of a tune. If we use the number of tonics as a proxy for how strongly established the key is, then we find that this interaction is the result of the key being that much more firmly indicated, which increases the impact tones that indicate major and minor. Since which tones sound major vs. minor is entirely determined by the tonic, it is unsurprising to find that the more strongly established a tonic is, the more firmly major or minor a sequence of tones sounds. If a listener were unsure about the key, one might expect the percept of major vs. minor to be weakened. Conversely, to the degree that a listener is certain of the key, we expect the percept of major vs. minor to be strong.

However, we also find that this interaction is not entirely symmetric with regard to major and minor. For both listeners, the values of the parameters ρ_k conform to a convex function as k increases from 1 to 8. It should be noted that the parameters ρ_k were forced to sum to zero; thus, this monotonic function was forced to assign negative (positive) values to ρ_k for low (high) values of k. What is clear, however, is that the sharpest variation in sensitivity to target notes occurs for strongly negative values of interaction strength ξ_j (Eq. 8). If we assume that the mechanism underlying these effects is activated most strongly in the region of its domain showing the steepest change, we are led to conclude that it is specifically the elements of the tone-scramble that promote a "minor" response to which the listener is most sharply sensitized by increasing tonic strength. This may indicate that there is a default to the major key when the tonic is established in the absence of evidence of either major or minor. When there is a strong key established, there does seem to be an extra boost in the major direction when the sequence is very major, but this boost caps out at a certain point. When there is a very strongly established tonic and the sequence is very minor, there is a very large boost and it does not cap out.

Though we do not know the true mean of the interaction parameters, as they were necessarily forced to sum to zero, it may be the case that there really is no major impact of high levels of major interaction. If the whole function is shifted down a little (and the criterion shifted down with it as well), we really only get an impact of minor interaction. Though this shift is arbitrary, we cannot say the same for shifting the interaction function upward, as the interaction function does not level out in the minor direction.

More evidence for the major mode being the default can be found in the tone functions, F and F_{last}. The tonics (G_5 and G_6) are either major or neutral in both of our participants. This may also suggest that this system defaults to the major mode.

Summary

This project has investigated (in two experienced listeners) the factors that control whether brief, proto-musical stimuli (called tone-scrambles) comprising a rapid random sequences of tones from the equal-tempered scale are judged to be major vs. minor. In all stimuli, the tonic was G, and the task was to judge whether the stimulus contained major target notes (major 3rds (i.e., B's) and/or major 6ths (i.e., E's)) or minor target notes (minor

3rd's (B♭'s) and/or minor 6th's (E♭'s)). The experiment was expressly designed so that the number of G's in different stimuli varied strongly across trials. This feature of the design enabled us to analyze the interaction between variations in the strength with which the tonic was established in a given stimulus and the other cues in the stimulus promoting a "major" vs. a "minor" response.

The data were described by a model that enabled us to disentangle (1) the influence exerted on the listener by each of the 13 notes *n* (ranging from G_5 to G_6) that occurred in these stimuli (a) when the note *n* occurred in a location other than the final location of the tone-scramble as well as (b) when the note *n* occurred in the final location; (2) the sensitivity of the listener to tones occurring at different temporal locations in the tone-scramble, and (3) the interaction between the strength with which the tonic was established in a given tone-scramble and the signal (as reflected by the sum of major vs. minor influences) in the tone-scramble.

Both listeners showed different patterns of sensitivity to notes occurring at the end of a tone-scramble than they did to notes occurring at non-final locations in the tone-scramble. For tones other than the final one, it was primarily the target notes that exerted the strongest influence (although slight idiosyncratic deviations from this pattern were shown by both listeners). By contrast, for both listeners, notes occurring at the final location tended to induce "minor" responses if they failed to provide a stable resolution to the tone sequence.

Both listeners showed a strong interaction effect. Sensitivity to (both major and minor) target notes was magnified by increased tonic strength. This effect, however, was stronger for minor than for major target notes suggesting that when a tonic is established to determine a key, the process that operates to sense the scale may default to major; under this hypothesis, the observed asymmetry might reflect the operation of a secondary process sensitive to minor cues in the presence of a strongly established tonic that can override this default assumption.

Conclusion

In this dissertation, I measured the sensitivity of listeners making many different kinds of judgments about fast randomly ordered sequences of musical tones. I found that scale degree 6 impacts listeners in almost identically to scale degree 3 with respect to major and minor. When participants are asked to make similar judgments about other sets of tones, though they do worse, they often do proportionately worse, implying there is only a single mechanism regulating these judgments, or that there is a single scaling factor that affects all of the mechanisms involved. When expert participants are asked to judge all combinations of tones, we find that there are likely 3 mechanisms involved in making these judgments. While it would seem that there is a mechanism distinguishing major and minor and a mechanism distinguishing consonance and dissonance, it is unclear what the third mechanism might be doing. We also asked listeners to make judgments of major and minor using stimuli with many other notes included to see how they impacted the listeners' judgment. We found that there is indeed an interaction between the major vs. minor judgments and the presence of the tonic, as well as a smaller impact of tones not relevant to the judgment being made.

References

Chew, E. (2013). *Mathematical and Computational Modeling of Tonality: Theory and Applications* (Vol. 204). Springer Science & Business Media.

Chubb, C., Dickson, C. A., Dean, T., Fagan, C., Mann, D. S., Wright, C. E., ... & Kowalsky, E. (2013). Bimodal distribution of performance in discriminating major/minor modes. *The Journal of the Acoustical Society of America*, *134*(4), 3067-3078.

Crowder, R. G. (1984). Perception of the major/minor distinction: I. Historical and theoretical foundations. *Psychomusicology: A Journal of Research in Music Cognition*, *4*(1-2), 3.

Crowder, R. G. (1985a). Perception of the major/minor distinction: II. Experimental investigations. *Psychomusicology: A Journal of Research in Music Cognition*, *5*(1-2), 3.

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

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

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

Halpern, A. R. (1984). Perception of structure in novel music. *Memory & cognition*, *12*(2), 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*, 335-355.

Heinlein, C. P. (1928). The affective characters of the major and minor modes in music. *Journal of comparative Psychology*, *8*(2), 101.

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

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

Kastner, M. P., & Crowder, R. G. (1990). Perception of the major/minor distinction: IV. Emotional connotations in young children. *Music Perception*, 189-201.

Krumhansl, C. L. (1979). The psychological representation of musical pitch in a tonal context. *Cognitive Psychology*, *11*(3), 346-374.

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.

Lerdahl, F. (1988). Tonal pitch space. *Music Perception*, 315-349.

Shepard, R. N. (1982). Geometrical approximations to the structure of musical pitch. *Psychological Review*, *89*(4), 305.

Wilks S. S. (1944). Mathematical statistics. Princeton: Princeton University Press.