The Effects of Timbre on Harmonic Interval Tuning and Perception

by

Na Eun Kim

Music Technology Area, Department of Music Research
Schulich School of Music
McGill University, Montreal
December 2017

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Master of Arts

Abstract

In Western music, temperament grants ease of performance and listening especially when it comes to harmony, whose multiplicity is built upon instrument intonation. However, our fluid perception allows other aspects of music such as timbre, blend and sensory dissonance to affect intonation. This thesis investigates how musicians perceive and compensate for the interacting effects of timbre, blend and sensory dissonance when tuning and rating harmonic intervals. The first two experiments involved an interval-tuning task and the third experiment involved a perceptual rating task for the unison, minor second, major third, tritone, perfect fifth, major sixth, minor seventh and octave. A different sample of twenty musically trained subjects participated in each experiment. In Experiment 1, participants tuned the upper note of an isolated harmonic dyad. Six timbre pairs comprised of the harpsichord, piano, clarinet and flute were chosen to reflect low to high blend. In Experiment 2, participants tuned the lower and upper note of an isolated harmonic dyad. Stimuli consisted of twelve pairs based on combinations of the harpsichord, trumpet, vibraphone and flute, chosen based on brightness and playing mechanism. In Experiment 3, participants rated all of the stimuli from Experiments 1 and 2 on three continuous scales representing subjective measures of auditory roughness, blend and pleasantness. Overall results showed lower tuning accuracy from equal temperament for musically consonant intervals regardless of pair, but higher tuning variability for musically dissonant intervals. Findings in relation to pairs with the flute and harpsichord supported previous research on increased distance in a timbre space suggesting decreased blend (Kendall and Carterette, 1993). Significant order differences in tuning were found mostly in relation to the pairs with the harpsichord as one of its instruments at consonant intervals. Musicians favoured interval contraction when tuning harpsichord pairs if the harpsichord was playing the upper note and interval expansion if it was playing the lower note, regardless of the instrument playing the note that was being tuned. Findings from Experiment 3 suggest that instrument assignment to upper and lower notes does not affect perceptual judgments of the overall timbre of a harmonic interval, but does affect the tuning of such intervals. Roughness and pleasantness ratings depended on the inclusion of the trumpet (rough, unpleasant) or vibraphone (smooth, pleasant), whereas blend ratings depended on whether or not instrument pairs had similar or different playing mechanisms. These perceptual ratings were not correlated with the tuning deviations

found in Experiments 1 and 2. Further research with a wider variation in timbre combinations will need to be conducted to explore whether or not these tuning differences are particular to the harpsichord due to its overall timbre, or due to a combination of its timbral properties that can be generalized to other instruments.

Résumé

Dans la musique occidentale, le tempérament musical bénéfice à la performance et à l'écoute, particulièrement sur l'harmonie dont la multiplicité est fondée sur l'intonation d'instrument. Cependant, la perception est flexible qu'il permet l'intonation d'être influencée par des aspects de musique comme le timbre, la fusion perceptive et la dissonance. Ce mémoire examine la perception des musiciens sur l'interaction des effets du timbre, de la fusion et de la dissonance lorsqu'ils font l'accordage des intervalles harmoniques. Les deux premières expériences laboratoires comportaient d'une tâche d'accordage d'un intervalle tandis que la troisième comportait d'une tâche d'évaluation perceptive pour l'unisson, la seconde mineure, la tierce majeure, le triton, la quinte parfaite, la sixte majeure, la septième mineure et l'octave. Chaque expérience était composée d'un échantillon différent de vingt participants de grande expertise musicale. Dans l'Expérience 1, les participants ont accordé la note supérieure d'une dyade harmonique isolée (DHI). Six paires de timbres (PT) composées du clavecin, du piano, de la clarinette et de la flûte ont été choisis pour refléter le mélange faible à l'élevé. Dans l'Expérience 2, les participants ont accordé les notes inférieure et supérieure d'une DHI. Les stimuli étaient composés de douze PTs combinant le clavecin, la trompette, le vibraphone et la flûte, choisis en fonction de la brillance sonore et du mécanisme de jeu. Dans l'Expérience 3, les participants ont évalué tous les stimuli des Expériences 1 et 2 sur trois échelles continues représentant les mesures subjectives de la rugosité sonore, de la fusion et du caractère agréable. Les résultats globaux montraient une précision d'accordage moindre par rapport au tempérament égal pour les intervalles musicaux consonants indépendamment des PTs, mais une variabilité d'accordage plus haute pour les intervalles dissonants. Les résultats concernant les PTs avec la flûte et le clavecin soutenaient la recherche précédente où l'augmentation de la distance dans un espace de timbres provoquait une fusion réduite (Kendall and Carterette, 1993). Des grandes différences de l'ordre des notes dans l'accordage étaient trouvées particulièrement sur les PTs comprenant le clavecin pour les intervalles consonants. Les musiciens ont préféré la contraction d'intervalle en accordant les paires avec clavecin s'il jouait la note supérieure et l'expansion d'intervalle s'il jouait la note inférieure, indépendamment de l'instrument qui jouait la note à accorder. Les résultats de l'Expérience 3 suggèrent que l'affectation des instruments aux notes supérieure et inférieure n'influence pas les jugements perceptifs du timbre d'un

intervalle harmonique, mais qu'elle influence l'accordage de tels intervalles. L'évaluation de la rugosité et du caractère agréable dépendaient de l'inclusion de la trompette (rugueuse et désagréable) ou du vibraphone (lisse et agréable), tandis que les évaluations de fusion dépendaient de la présence ou l'absence des mécanismes de jeu similaires ou différents dans les paires d'instruments. Ces évaluations perceptives n'étaient pas corrélées avec les déviations d'accordage dans les Expériences 1 et 2. Des recherches supplémentaires sont nécessaires avec une variation plus large des combinaisons de timbre pour examiner si ces différences d'accordage sont spécifiques au clavecin en raison de son timbre global ou bien en raison de la combinaison des caractéristiques de son timbre qui peuvent être généralisées à d'autres instruments.

Acknowledgements

I would like to express my gratitude toward my supervisor, Stephen McAdams, for willingly taking me under his wing during my graduate studies. He has been a patient and solid pillar of support in all aspects, whose wealth of knowledge never fails to amaze me. I would like to give a special shout out to Bennett for sharing his technical expertise, to Cecilia Taher and Jason Noble for their moral and practical support, and to Manda Fischer for being a constant source of encouragement in the lab. I would also like to thank all of the wonderful members of the Music Perception and Cognition Lab (MPCL) for consistently taking the time to pilot my experiments and provide statistical advice, namely: Kai Siedenberg, Meghan Goodchild, Etienne Thoret, Eddy Kazazis, Ian Lou, Moe Touizrar, Kit Soden, Tanor Bonin, Lena Heng, Hannah Darroch, Behrad Madahi and Pierre Gianferrera. I am especially thankful to Negin Abaeian for her emotional and academic assistance, and to Quim Llimona, Marina Azon and Olafur Bogason for their laughter and humorous tales. Finally, I would like to thank all of the friends that I made during my time in Montreal, my dear family and friends back in Toronto, and God above all for their endless love and motivational inspiration.

Author Contributions

As the sole author, I was responsible for designing and executing the experiments, data analysis and interpretation, and writing and editing this thesis. My thesis advisor and supervisor, Stephen McAdams, provided financial support, laboratory access and equipment, and additional guidance in terms of experimental design, statistical analysis and data interpretation. MPCL's Technical Manager, Bennett K. Smith, assisted with the preparation and modification of the stimuli as well as the creation of all loudness matching and experimental programs.

Contents

List of Figures

List of Listings

List of Tables

List of Acronyms

RMS	Root mean square
dB	Decibel
SPL	Sound pressure level
ANOVA	Analysis of variance
12TET	12-tone equal temperament
GLM	General Linear Model
SD	Standard deviation
CI	Confidence interval
P1	Unison
m2	Minor second
M3	Major third
A4	Tritone
P5	Perfect fifth
M6	Major sixth
m7	Minor seventh
P8	Octave
HC	Harpsichord
PN	Piano
CL	Clarinet
FL	Flute
TR	Trumpet
VB	Vibraphone

Chapter 1

Introduction

1.1 Timbre and Tuning in Tonal Harmony

Timbre is a multidimensional construct correlated with certain physical properties of sound such as attack time, spectral centroid, spectral deviation and spectral flux (McAdams et al., 1995). In music, it is what helps us distinguish and classify the acoustic quality of different instruments and to describe the characteristics of a sound event. Intonation cannot readily be separated from timbre in real-world sounds, because the spectra of many musical instruments do not always correspond to a perfect harmonic series and can contain either stretched or compressed partials (Hasegawa, 2009). This is why despite the fact that musically consonant intervals such as the octave should produce little to no roughness at simple frequency ratios (Pressnitzer et al., 2000), certain combinations of instruments playing in harmony can still sound “off” to the listener even if they are individually in tune. Indeed, Kopiez (2003) found that intonation is affected by a combination of “compositional features, the acoustics of the particular musical instrument and deviation patterns in specific intervals” (p. 383). The physiological source of such incongruity between what should theoretically sound consonant to what is perceptually dissonant is largely based on the detection of auditory roughness, a phenomenon arising from the beating from rapid amplitude fluctuations between adjacent partials in a sound (Plomp & Levelt, 1965; Pressnitzer et al., 2000). Since sensory dissonance depends on timbre (Kameoka & Kuriyagawa, 1969a, 1969b; Mashinter, 2006) and auditory roughness results as a by-product of the instrumental timbre combinations in an orchestrated sound event; the perception of roughness is not limited to what is traditionally considered dissonant in music theory.

1.1.1 Musical versus Sensory Consonance

The concepts of consonance and dissonance in music have frustrated and fascinated researchers since before the development of polyphony (a composition style that incorporates two or more independent melodic structures played simultaneously). According to James Tenney (1988), the "consonance/dissonance concept (or CDC)" (p.3) has "been used, historically, in at least *five different ways*" (p. 4). Tenney proposes that distinctions must be made between conceptions, theories, aesthetic attitudes and practical uses of consonance and dissonance, through which different aspects of this "acoustical/musical/perceptual *phenomenon*" (p. 5) can manifest themselves. Terhardt (1984) integrates the psychoacoustic evaluation of sensory consonance and the musical experience of harmony into a two-part definition of consonance/dissonance. "*Sensory consonance*…represents the graded absence of annoying factors and is not confined to musical sounds" whereas "*harmony*…represents the typical, music-specific principles of tonal affinity, compatibility, and fundamental-note relations (root)" (p. 276). The psychoacoustic model for consonance defines it as the absence of "acoustically dissonant factors" (Hutchinson & Knopoff, 1978, p. 2) or auditory roughness, thereby defining the existence of auditory roughness as sensory dissonance. Although consonance and dissonance may very well be a multidimensional construct composed of an amalgamation of arithmetic, physics, physiology, psychology and culture rooted in musical context (Di Stefano & Bertolaso, 2014), the current body of empirical research has focused on the interactions between the well-known dimensions of musical consonance (dissonance) based on theoretical ideals and sensory consonance (dissonance) based on psychoacoustics and perception.

Hall and Hess (1984) found that musicians' "*judgments represent some combination of beat-rate detection and semitone-fraction awareness*" (p. 190), and suggested that tuning perception is based on the knowledge of interval sizes as well as the ability to hear auditory roughness. Johnson-Laird, Kang and Leong (2012) found that musically consonant chords are perceived as less dissonant in a tonal sequence compared to a random sequence and proposed "a dual-process theory that embeds roughness within tonal principles" (p. 19). Building on preceding psychoacoustic models of auditory roughness, Sethares (1993, 1994) created alternative, timbre-specific tuning systems based on the local consonance and dissonance curves of a sound's spectrum, allowing any timbre to achieve optimal sensory consonance. Sethares (1998) posed that "almost any interval can be made dissonant or consonant by proper sculpting of timbre" (p. xii). Based on the findings of Vos (1982), Krumhansl (1991) indicated that although "interference between harmonics is the major influence on judgments of tuning…additional, nonpsychoacoustic factors or strategies also operate" (p. 281). This is supported by studies looking at the neural correlates of sensory consonance in the brain, which showed that there is more going on than merely the perception of roughness or beating in a harmonic sound when it comes to recognizing sensory consonance and dissonance (Bidelman & Heinz, 2011; Cousineau et al., 2012). Much of the research conducted in relation to intonation focus on isolated sequential tones or melodious sequences (Hutchins et al., 2012; Rakowski, 1990; Schellenberg, 2002; Swaffield, 1974; Wapnick & Freeman, 1980), but music is more often composed for and performed by a group of instruments. As such, tonal fusion becomes another important factor to consider when studying sensory dissonance, because "grouping processes…influence our perception of the qualities of source images (including pitch)" (McAdams, 1984, p. 303). Although quite a few studies have looked at various instrument intonation and tuning tendencies at harmonic intervals for different timbres (Karrick, 1998; McDermott et al., 2010; Platt & Racine, 1985; Geringer & Witt, 1985; Geringer & Worthy, 1999; Worthy, 2000), not all have studied how these intonation tendencies are affected by the interaction of timbre and tonal fusion (blend).

1.2 Timbre, Blend and Sensory Dissonance

Brues (1927) studied tonal fusion in harmonic intervals and found that the presence of auditory roughness reduces fusion, or "unitariness." (p. 626). Moore, Glasberg and Peters (1986) highlighted the importance of harmonicity to fusion, showing that mistuned partials in a complex tone stand out separately. DeWitt and Crowder (1987) found that consonant intervals promote tonal fusion and may "represent harmonics resulting from a single fundamental as timbres rather than chords" (p. 73). Research also suggests that increased distance within a timbre space will decrease blend between two instruments (Kendall & Carterette, 1993), whereas Sandell (1995) found that decreased composite brightness (the spectral centroid of the composite spectrum), decreased average brightness (the difference between the centroids of paired instruments) and increased onset synchrony between two instrumental tones will increase blend. Shields and Kendall (2004) found that timbre (spectral centroid) and interval size affect blend and dissonance ratings; brightness has the most influence on dissonance ratings for musically dissonant intervals, and the timbre of the upper voice in intervals is more salient than the lower voice.

1.3 Experiment Overview

The three experiments carried out for this thesis were prompted by the phenomenon in which instrumental timbres can sound harmonious when paired with certain timbres and out-of-tune when paired with others at musically consonant intervals, even when there is no perceptible difference in intonation. Since there seems to be more to sensory dissonance than just the psychoacoustic perception of roughness, how exactly does instrumental timbre affect the way people perceive sensory dissonance and blend in harmonic intervals? How much do these perceptions affect tuning, and are there consistent differences between diverse timbre combinations at various musical intervals? These experiments were meant to explore the relationship between musical (theoretical) and perceptual (sensory) consonance/dissonance, how tuning is used to reconcile their differences and how timbre and instrumental blend affect the way we perceive and tune harmonic intervals in Western tonal music. All three studies were certified for ethical compliance by the McGill University Research Ethics Board II prior to experimentation.

1.3.1 Apparatus

All three experiments were run using the same PsiExp interface (Smith, 1995) on a Mac Pro 5 computer running OS 10.6.8 (Apple Computer, Inc., Cupertino, CA) inside an IAC model 120act-3 double-walled audiometric booth (IAC Acoustics, Bronx, NY). Stimuli were presented over Sennheiser HD 280 Pro headphones (Sennheiser Electronic GmbH, Wedemark, Germany) and amplified through a Grace Design m904 monitor (Grace Digital Audio, San Diego, CA) at a relative level of 60 dB. The dB SPL range was obtained by first calculating the root mean squares (RMS) in dB using Sound eXchange of all of the combined stimulus pairs in the experiment, and then measuring the physical sound levels of the sound files with some of the lowest (e.g., –26 to –25) and highest RMS values (e.g., –11 to –9) through the headphones using a Brüel & Kjær Type 2205 sound-level meter (A-weighting) placed at the level of the listener's ears (Bruel & Kjær, Nærum, Denmark). The physical sound levels of all stimuli ranged from 47 to 67 dB SPL. One explanation for this gap in dB SPL stems from the fact that all loudness-matching procedures were conducted pre-experiment, whereas sound levels were measured post-experiment. Loudness matching was based on the median subjective perceptual ratings of loudness across participants, which resulted in percussive pairs with strong attacks (i.e., HC, PN) being matched at a quieter dB level compared to sustained pairs (i.e., FL, CL). The exception was the vibraphone, a percussive instrument, being matched at a higher dB level due to the fact that it sounded very much like a pure tone and seemed perceptually softer than the other instruments. The vibraphone samples used, however, were recorded with a soft mallet, which likely contributed to its muted resonance.

Chapter 2

Experiment 1: Timbre and Harmonic Interval Tuning 1

This first experiment investigated whether or not asymmetries in tuning would exist in timbre pairs chosen based on their distance in a timbre space to reflect different levels of blend and sensory dissonance. The existing body of literature suggests the following: 1) increased distance between a pair of two different instruments in Kendall & Carterette's (1991) timbre space predicts decreased blend in a harmonic dyad (Kendall & Carterette, 1993), 2) increased sensory dissonance is associated with decreased blend (DeWitt & Crowder, 1987; Shields & Kendall, 2004) and 3) musically dissonant intervals are also associated with decreased blend (DeWitt & Crowder, 1987) and greater influence of timbre (Shields & Kendall, 2004). With these points in mind, we hypothesized that the associated levels of blend and sensory dissonance with the levels of distance between two instruments in a timbre pair would increasingly affect tuning accuracy, especially at musically dissonant intervals. Six timbre pairings consisting of the harpsichord, piano, clarinet and flute (HC/PN, PN/HC, CL/FL, FL/CL, FL/HC, HC/FL), selected on the basis of their distances within the Lakatos (2000) timbre space, were played at eight different musical intervals (the unison, minor second, major third, tritone, perfect fifth, major sixth, minor seventh and octave, notated as P1, m2, M3, A4, P5, M6, m7 and P8, respectively). Under our assumptions, the timbre pairs with the furthest distance between their instruments (FL/HC and HC/FL) and played at musically dissonant intervals (m2, A4, m7) would show the highest tuning variability and difficulty and lowest tuning accuracy, presumably due to their low degree of blend and high degree of sensory dissonance.

We were also interested in whether or not participants would find FL/HC and HC/FL more difficult to tune than the other pairs even at musically consonant intervals due to their low presumed degree of blend and high sensory dissonance, despite Shields and Kendall's (2004) finding that suggests timbre perceptually affects musically dissonant intervals only. Finally, because Shields and Kendall (2004) found that the timbre of the upper voice was more perceptually salient, we expected some differences to occur between HC/PN & PN/HC, CL/FL & FL/CL and FL/HC & HC/FL due to the position of the instruments. Greater mean deviations from 0 would suggest lower tuning accuracy with respect to 12-tone equal temperament (12TET), whereas tuning variability would suggest higher tuning difficulty.

2.1 Method

2.1.1 Participants

Twenty musicians recruited from McGill University participated in this experiment (mean age = 22.2 years, SD = 2.65). All but two of the participants had over ten years of experience (mean years = 14.8, SD = 3.78) and at least 1 year of university-level ear training (mean years = 7.1, SD = 4.15). 80% of the musicians who participated in the study were multi-instrumentalists. In addition, 15% had conducting experience and 10% had composing or arranging experience. Out of all instruments mentioned, eleven musicians had experience playing woodwinds (flute, clarinet, bassoon, oboe, saxophone), sixteen had experience playing keyboards (piano), four had experience playing percussion (drums), nine had experience playing stringed instruments (guitars, violin, cello, bass), five had experience playing brass (trumpet, French horn, euphonium) and seven had experience performing with vocals.

All participants met the required threshold of 20 dB HL on a pure-tone audiometric test with octave-spaced frequencies from 125 Hz to 8 kHz (ISO 389–8, 2004; Martin & Champlin, 2000) before proceeding to the experiment. All participants signed a written consent form before following through to the experimental session and were debriefed after completing all steps of the experiment. Participants were paid $10 CAD for roughly 30 minutes to an hour of their time.

2.1.2 Stimuli

Six pairs of sounds with different timbres were chosen based on their distance in the Lakatos (2000) combined space. The first two timbre pairs consisted of the harpsichord and piano (HC/PN, PN/HC) with a close distance of 0.14 on a scale from 0 to 1. The third and fourth pairs consisted of the clarinet and flute (CL/FL, FL/CL) with a medium distance of around 0.22. The final two pairs consisted of the flute and harpsichord (FL/HC, HC/FL) with a large distance of around 0.52. The flute, harpsichord and clarinet sounds were taken from the Vienna Symphonic Library (https://vsl.co.at), whereas the piano sounds were taken from the Kontakt 3 Library (August Förster collection). The range of notes spanned an octave from C4 to C5 in order to make up eight intervals; the notes used were C4, C#4, E4, F#4, G4, A4, A#4 and C5. All pitches for each instrument were verified to match or closely match the fundamental frequencies of 12-TET with A4 at 440Hz using a MATLAB code that calculated the highest peak frequency (see Listing. 2.1). Version 2.1.2 of Audacity®, a recording and editing software, was used to find the fundamental frequency with a Hamming window at 8192-point resolution if the MATLAB code's highest peak frequency was not the fundamental.

```
clear all;
close all;

[y, fs] = audioread('HC_60.wav') ;
FFT_ = fft(y) ;
L = length(y) ;

f = fs*(0:(L/2))/L;

plot(f(1:end-1),abs(FFT_(1:end/2)));

[m,i] = max(abs(FFT_(1:end/2))) ;

f(i);
```

Listing. 2.1: MATLAB code used for calculating the highest peak frequency of HC at C4.

Unfortunately, the August Förster collection was missing the piano notes of C#4, D#4, F#4, G#4 and A4, which required transposition from the nearest existing semitone using a high-quality resampling algorithm (Sound eXchange, http://sox.sourceforge.net) for pitch-shifting. Sound eXchange (SoX) was also implemented in the experiment code, using linear phase resampling and a resampling bandwidth of 95% (the amount of audio frequency bands that are preserved in the signal). Resampling filters can potentially cause transient echo artefacts in highly percussive signals, and the pre-echoes of such percussive signals could be perceptually noticeable. However, when observing the impulses of a note that was adjusted +150 and -150 cents (the maximum possible directional adjustments of a pitch-shifted sound sample used in the study), it was found that the SoX algorithm adjusts the phases so that only post-echoes remain, lasting for around 1.3ms and attenuated by at least 30dB. These post-echoes are masked by the attack of the sound sample and are not perceptible.

Each stimulus lasted for one second. The physical onsets for some of the stimuli were also adjusted to increase perceived onset synchrony, and a loudness-matching experiment was conducted prior to the experiment to equalize the stimuli. A preliminary program was created to fine-tune the onset delays and amplitudes of the different timbre pairs at each interval. These parameters were recorded among three people and averaged. Ten musicians who were excluded from the experimental study participated in the loudness-matching experiment, in which the loudness of all the stimuli was matched to the harpsichord at C4. The harpsichord was chosen as the comparison tone because it was perceptually the loudest among the four instruments. Median loudness values were calculated across all ten participants and were used to adjust stimulus levels to achieve loudness equalization. The adjusted onset delay and amplitude parameters were included in the tuning experiment along with the median loudness values.

2.1.3 Experimental Design

The experiment was a 6x8 repeated-measures design consisting of two within-subjects factors: six timbre pairs based on three distances and matched according to relative position in the dyad (hereafter referred to as "order": HC/PN, PN/HC, CL/FL, FL/CL, FL/HC, HC/FL) and eight pitch intervals (P1, m2, M3, A4, P5, M6, m7 and P8). The instrument pairs were chosen to reflect close, medium and far distances in the Lakatos timbre space, which was determined only on the basis of D#4. The intervals were chosen specifically to have a mix of perfect consonant (P1, P5, P8), imperfect consonant (M3, M6) and dissonant (m2, A4, m7) intervals.

2.1.4 Procedure

Baseline trials, in which the same instrument played the upper and lower notes, were mixed in with the experimental trials, in which different instruments played the upper and lower notes. There were no baseline trials included for P1. The experiment was divided into three blocks, the first consisting of 8 practice trials and the second and third consisting of 38 randomized trials each, for a total of 84 trials. Following completion of the experiment, participants filled out an online questionnaire.

Participants were required to adjust the pitch of the upper note of a given harmonic interval (limited to half a semitone above and below, spanning ±50 cents around the zero adjustment mark) with the lower note of the dyad fixed at C4. They were instructed (see Appendix B, Experiment 1 Instructions, p. 76) to tune the upper note to make the target harmonic interval sound as "acoustically in-tune as possible" (also phrased as "perceptually in-tune" during explanations). Participants were asked to tune subjectively, rather than follow the specific interval sizes of existing temperaments (e.g., 12-tone equal temperament, just intonation). Each trial displayed a screen that showed the pitch interval participants were required to adjust, with the given pitch interval playing automatically over the headphones. The initial sound programmed the upper note to be either sharper or flatter than its original temperament by a random number of cents. Participants were free to tune the upper note using small (3 cents) and/or big (10 cents) steps to get as close as possible to the most "acoustically in-tune" sound for that interval.

The interface layout consisted of written instructions at the top of the screen, separately clickable buttons for small steps (on the left) and big steps (on the right) below the instructions, separately clickable buttons for '<' (change to lower pitch), play and '>' (change to higher pitch) in the middle of the screen, and a button for the next trial at the very bottom.

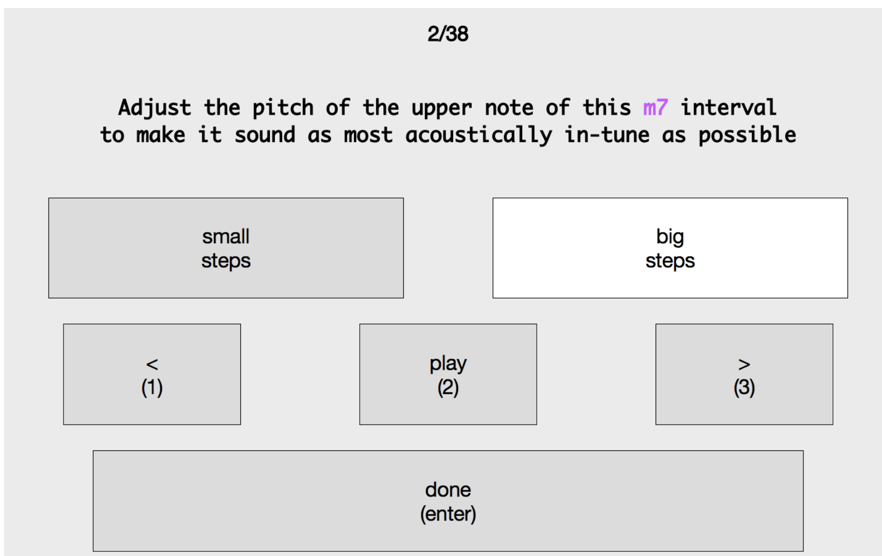

Fig. 2.1: A screenshot of the interface from Experiment 1.

Participants first clicked on the step interval before using the left and right arrows to fine-tune the sound. The left arrow signified a step down, which made the sound either 3 or 10 cents flatter, and participants had the option to press the '1' key instead of mouse-clicking the button. The right arrow signified a step up, which made the sound either 3 or 10 cents sharper, and participants had the option to press the '3' key instead. The program automatically played the adjusted sound after every edit, but the 'play' button was included so that participants could listen to the edited result again. This button was accessible with either a mouse-click or the ‘2’ key. Once participants were satisfied with their tuning, they were required to either click the ‘done’ button or press ‘enter’ to move on to the next trial.

2.2 Results

All analyses were measured against one dependent variable, Cent_Deviation (tuning deviation from zero, which is equivalent to 12-TET). Analyses involved two independent variables (Instrument and Interval for baseline trials, and Pair and Interval for experimental trials). The data were analyzed in the following ways: 1) an analysis of baseline means using a 4 Instrument x 7 Interval (excluding interval P1) General Linear Model (GLM) repeated-measures analysis of variance (ANOVA) in SPSS, 2) an analysis of experimental means using a 6 Pair x 8 Interval GLM for Repeated Measures ANOVA in SPSS, 3) a series of paired-samples t-tests to compare baseline and experimental means at the Bonferroni-adjusted alpha-level of $\alpha = .05/42 = .0012$ and 4) an analysis of the 95% confidence intervals around baseline and experimental means to check for tuning accuracy and variability.

2.2.1 Analysis of Baseline Means

In the baseline analysis, sphericity was not violated for Instrument, $\chi^2(5) = 10.03$, $p = .08$, but was violated for Interval, $\chi^2(20) = 52.26$, $p < .001$ and for the Instrument*Interval interaction, $\chi^2(170) = 254.56$, $p < .001$. Greenhouse-Geisser corrections were used for Interval, $\varepsilon = .55$ and Instrument*Interval, $\varepsilon = .44$. The main effect of Instrument and Interval were significant, $F(3, 57) = 3.30$, $p = 0.27$ and $F(3.29, 62.58) = 3.24$, $p = .02$, respectively. These main effects suggest differences in tuning variability between instruments as well as between intervals in same-instrument conditions. The Instrument*Interval interaction was also significant, $F(7.92, 150.53) = 2.45$, $p = .02$.

Bonferroni-adjusted pairwise comparisons showed significant differences only in relation to HC/HC at intervals m2, P5 and P8. At m2, HC/HC was tuned flatter than PN/PN and CL/CL. At P5, HC/HC was tuned sharper than CL/CL and FL/FL. At P8, HC/HC was tuned flatter than PN/PN. This may suggest decreased tuning accuracy for the harpsichord compared to the other instruments. The other instruments were not tuned with significant differences between each other at any interval.

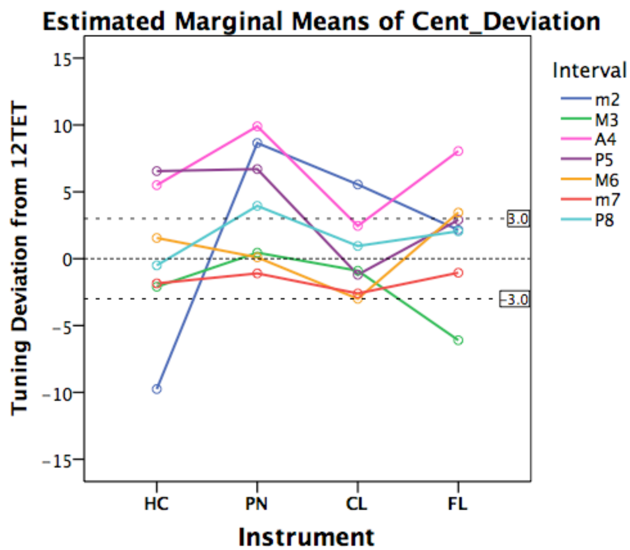

Fig. 2.2: Baseline mean cent deviations for Interval as a function of Instrument. Dashed lines show the smallest tuning change increment of 3 cents.

An investigation of baseline means within instruments across intervals showed that the harpsichord had the most significant differences within itself, whereas the clarinet had no significant differences at all. For HC/HC, m2 was tuned significantly lower than A4 and P5. P5 was tuned significantly higher than M3 and P8. For PN/PN, M3 was tuned significantly lower than A4 and P5. For FL/FL, M3 was tuned significantly lower than A4. The harpsichord, piano and flute seem to display lower tuning accuracy at certain intervals. The lack of significant differences for the clarinet suggests that participants were the most accurate when tuning the clarinet and were the most inaccurate when tuning the harpsichord.

2.2.2 Analysis of Experimental Means

For the experimental analysis, sphericity was not violated for Pair; $\chi^2(14) = 12.86$, $p = .54$. However, it was violated for Interval; $\chi^2(27) = 65.56$, $p < .001$, so the Greenhouse-Geisser correction was used, $\varepsilon = .48$. The Pair*Interval interaction did not produce a sphericity value in SPSS, so the Greenhouse-Geisser correction was used as well, $\varepsilon = .3$. The main effects of Pair and Interval were significant; $F(5, 95) = 12.06$, $p < .001$ and $F(3.37, 63.96) = 2.69$, $p = .05$, respectively. The Pair*Interval interaction was also significant, $F(10.53, 200.14) = 3.01$, $p =$

.001. These results suggest significant differences in tuning accuracy between timbre pairs and between intervals in different-instrument conditions, and also suggest differences in timbre pairs across certain intervals. Differences between timbre pairs overall as well as differences between timbre pairs at certain intervals are telling of the importance of order in the upper and lower notes of each matched instrument pair. The main effect of Pair was analyzed to see if there were globally significant differences between matched pairs. Only FL/HC and HC/FL were significantly different, with FL/HC tuned flatter than HC/FL, suggesting that instrument order (i.e., their position as either the lower or upper note) was especially important for these pairs.

Analysis of Bonferroni-adjusted post-hoc pairwise comparisons for the interaction effect between Pair and Interval showed that at interval P1, HC/PN was tuned significantly higher than PN/HC, whereas FL/HC was tuned significantly lower than PN/HC and HC/FL (it was also tuned significantly lower than every other pair). At interval P5, CL/FL was tuned significantly lower than HC/FL. At interval P8, PN/HC was tuned significantly lower than HC/PN. HC/FL was tuned significantly higher than CL/FL and FL/HC. Similar to baseline analyses, significant results were obtained exclusively at consonant intervals. There appears to be a flattening effect when the harpsichord is tuned as the upper note at musically consonant intervals, especially when paired with the flute. When tuning the flute at P5 and P8, participants tuned the flute sharper when the harpsichord was the lower note compared to when the clarinet was the lower note. There were no significant differences between matched pairs CL/FL and FL/CL.

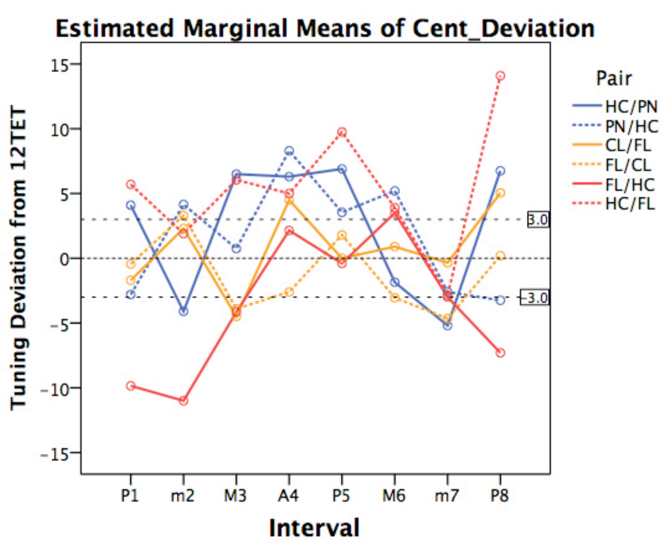

Fig. 2.3. Experimental mean cent deviations as a function of Interval for each Pair.

2.2.3 Comparison of Baseline and Experimental Means

Forty-two Bonferroni-corrected paired-samples t-tests were run in SPSS, in which baseline measures and instrumental pairs with the same instrument tuned as the upper note were compared (e.g., Baseline measures for HC were only compared with PN/HC and FL/HC, where HC was the instrument being tuned). Only the p-values at $\alpha = .0012$ or below were declared significant.

At P8, HC/HC was tuned significantly higher than FL/HC. This result provides support for the flattening effect that occurs for the FL/HC pairing at the octave.

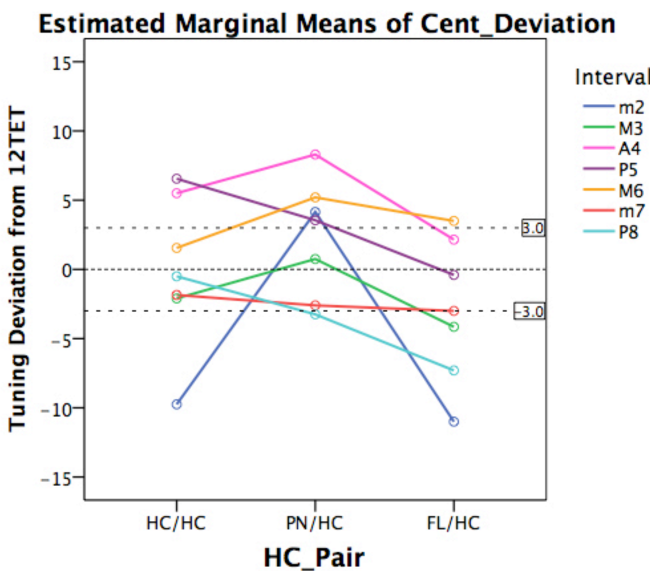

Fig. 2.4. Baseline and experimental mean cent deviations for Harpsichord.

At m2, PN/PN was tuned significantly higher than HC/PN. The lack of significant results was expected, as HC/PN was chosen for its relatively close distance parameter and associated with higher blend and lower sensory dissonance.

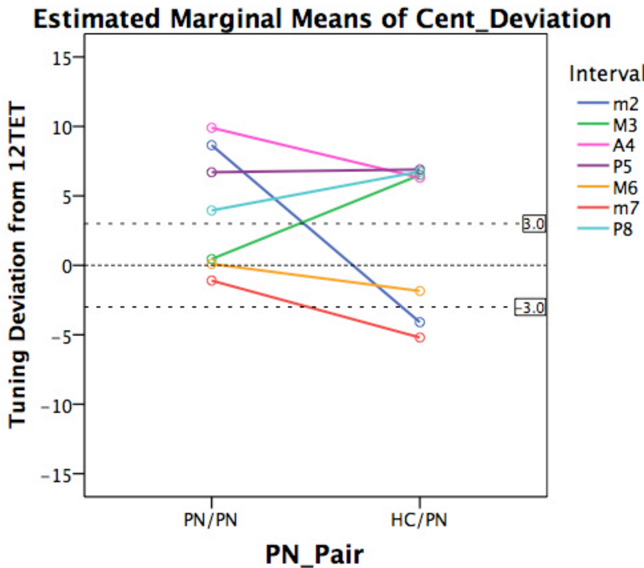

Fig. 2.5. Baseline and experimental mean cent deviations for Piano.

At P8, FL/FL was tuned significantly lower than HC/FL. It is possible that the flute as the lower note is responsible for the flattening effect that occurs for FL/HC at consonant intervals.

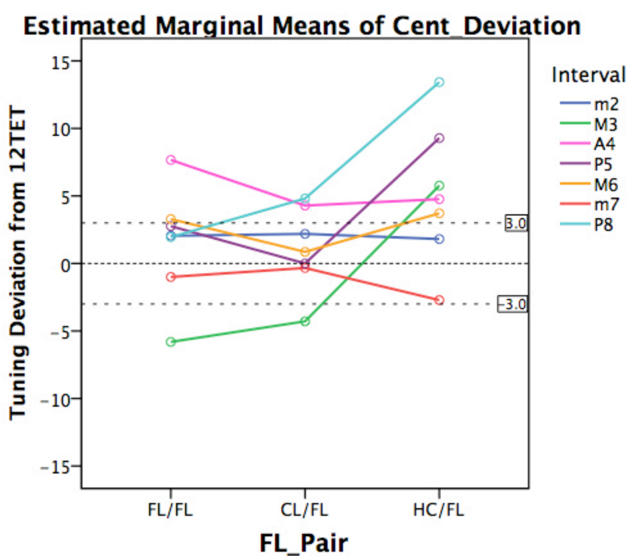

Fig. 2.6. Baseline and experimental mean cent deviations for Flute.

Significant differences occurred exclusively at intervals m2 and P8, and Clarinet was the only instrument with no difference between any baseline and experimental measures when it was tuned as the upper note.

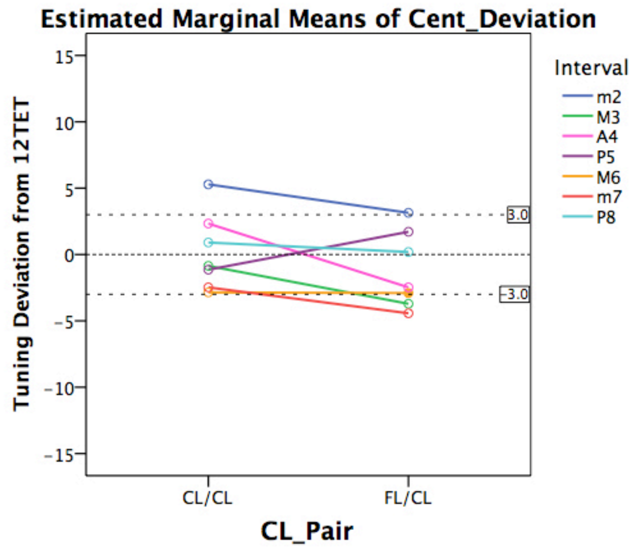

Fig. 2.7. Baseline and experimental mean cent deviations for Clarinet.

2.2.4 Analysis of 95% Confidence Intervals around Baseline and Experimental Means

The 95% confidence intervals for both ANOVA datasets were analyzed to see if tunings were significantly different from zero (in other words, from 12TET). If zero was included within the CI bounds, it was assumed that the mean cent deviation was not significantly different from 12TET tuning, which could indicate higher tuning accuracy. If zero was not included within the upper and lower bounds of the confidence interval, the opposite was assumed.

For the family of baseline means, the Instrument*Interval interaction was studied to examine the tuning accuracy and variability for each instrument at each interval. Harpsichord at P5 was tuned sharp with a 95% CI of 5.1 to 8.0 cents. Piano was tuned sharp at A4 with a 95% CI of 4.4 to 15.5, sharp at P5 with a 95% CI of 3.2 to 10.3, and sharp at P8 with a 95% CI of 2.7 to 5.2. Flute was tuned sharp at P5 with a 95% CI of 1.3 to 4.5. There are several points of interest: significant deviations from 12TET happened at mostly consonant intervals (perfect fifth and octave), mostly occurred for the piano, did not occur at all for the clarinet and all were sharp in nature. Based on our estimations, significant deviations from 12TET assume lower tuning accuracy. Intuitively, it might make more sense for lower accuracy to occur at musically dissonant intervals compared to consonant intervals, based on the common conceptions of sensory dissonance in music and tuning. However, no significant deviations were found at dissonant intervals except for the piano at A4. Although these cases meet the requirements for being tuned differently from 12TET, tuning variability (the range of confidence intervals) for consonant intervals was lower compared to that of dissonant intervals. Lower tuning accuracy in relation to 12TET does not necessarily mean higher tuning difficulty, judging from the smaller deviation ranges that occurred at consonant intervals.

The Pair*Interval interaction was also examined for the family of experimental means. HC/PN was tuned sharp at P5 with a 95% CI of 3.8 to 10.0, and sharp at P8 with a 95% CI of 3.1 to 10.4. FL/HC was tuned flat at P1 with a 95% CI of –12.9 to –6.8, and flat at P8 with a 95% CI of –10.6 to –3.9. HC/FL was tuned sharp at P5 with a 95% CI of 5.4 to 14.2, and sharp at P8 with a 95% CI of 11.6 to 16.6. Greatly significant deviations happened at the most musically consonant intervals (unison, perfect fifth and octave) and CL/FL and FL/CL did not have means that significantly differed from 12TET. The only significantly flat deviations occurred exclusively for FL/HC, and all other significant deviations were sharp in relation to 12TET.

Overall, deviation ranges were larger and more variable for dissonant intervals compared to consonant intervals, but it is unclear from this analysis if timbre pairs FL/HC and HC/FL were the most difficult to tune at musically dissonant intervals. However, it does appear to show some support for the hypothesis that participants are generally less accurate when tuning FL/HC and HC/FL than the other timbre pairs at musically consonant intervals. Despite this, it is entirely possible that the presence of the harpsichord in these timbre pairs played a part, because all of the significant deviations were related to pairs that included the harpsichord as either the upper or lower note.

2.3 Discussion

The results provided partial support for our initial hypotheses. Interpretations of decreased tuning accuracy were drawn from greater deviations from zero, whereas interpretations of increased tuning variability were associated with greater 95% CI ranges and standard deviations. Although FL/HC and HC/FL did appear to be associated with lower tuning accuracy compared to PN/HC, CL/FL and FL/CL, it was only at the unison, perfect fifth and octave. Based on initial assumptions and data analyses, it appears that the decreased blend associated with greater distance within a timbre space affects tuning accuracy at musically consonant intervals, but not necessarily at musically dissonant intervals. Overall, the data do not statistically suggest that FL/HC and HC/FL were more difficult to tune than the other timbre pairs on the basis of its instrument combination. Regardless of pair, musically dissonant intervals did not show any statistically significant deviations from 12TET, which was surprising since Shields and Kendall (2004) found that timbre had the greatest influence at musically dissonant intervals. Although musically dissonant and imperfect consonant intervals did not show clear tuning inaccuracies in relation to 12TET, they had greater standard deviations and confidence interval ranges overall, which may be a reflection of tuning difficulty due to sensory dissonance. However, because sensory dissonance was not controlled for, per se, this is only a speculation.

The clarinet was the only instrument with no difference between any baseline and experimental measures, suggesting that it was no easier or harder to tune when it was simultaneously played with another instrument compared to when it was played with itself. Timbre pairs that included the clarinet comparatively showed the highest tuning accuracy with the lowest mean cent deviations from 12TET, despite having higher distance numbers than HC/PN and PN/HC. It appears that compared to HC/HC, FL/HC contracted the harmonic interval size at P8, and that compared to FL/FL, HC/FL expanded the harmonic interval size at P8. The timbral distances of the pairs FL/HC and HC/FL are equal, so the importance of the position of instruments is evident, at least at the octave. This trend is reminiscent of the results presented in Vurma et al.'s (2010) paper, where the trumpet and tenor (voice) sounds of a bright timbre were tuned 15 to 20 cents sharper at the unison than the viola sound of a duller timbre, although the tones were presented melodically in their study. In the context of our experiment, the addition of the flute may have dulled the composite brightness of FL/HC compared to HC/HC, and the addition of the harpsichord may have brightened the composite spectrum of HC/FL compared to FL/FL. Because participants had to tune only the upper note, the change in the timbre of the lower note clearly affected the results for these instrumental pairings, as the lower timbre at C4 was essentially the comparison note.

According to Sandell's (1995) findings, decreased composite brightness and decreased average brightness between two instrumental tones increase blend. In the case of FL/HC and HC/FL, it is possible that FL/HC's decreased composite brightness increased its blend, whereas HC/FL's increased composite brightness decreased its blend. If higher sensory dissonance is associated with decreased blend (DeWitt & Crowder, 1987; Shields & Kendall, 2004), then HC/FL should be comparatively more dissonant than FL/FL, and FL/HC should be comparatively less dissonant than HC/HC. However, these speculations based on instrument order only apply to the octave and cannot be generalized to the other intervals.

Judging from the main effect of Pair, flute and harpsichord pairings were the only pairs for which it was evident that instrument order was globally significant. FL/HC contracted the harmonic interval size at the octave (participants tuned HC flatter), whereas HC/FL expanded it (participants tuned FL sharper). This could be due to their relatively large distance in the timbre space, which is associated with decreased blend. All of the significant deviations from 12TET for experimental means were related to pairings that had harpsichord as either the upper or lower note. Harpsichord was also the most inaccurately tuned even for baseline measures, suggesting that its timbre more noticeably affected the results. It is possible that the majority of participants thought the harpsichord sounded more "acoustically in tune" when not following the 12TET standard, which was reflected in its inaccurate tuning results. The caveat here is that the harpsichord samples used double stops (two strings per note), which makes many of its higher frequencies more perceptible. This could have been a contributing factor to the way musicians tuned these harpsichord samples.

Interaction effects between Pair and Interval showed that, interestingly, FL/HC was always tuned flatter than the other pairs at the unison and octave, whereas HC/FL was always tuned sharper than other pairs at the unison, perfect fifth and octave. Like the baseline and experimental mean comparisons suggest, perhaps musicians like to tune the upper note of a harmonic interval sharper if the comparatively brighter harpsichord plays the lower note, and tune the harpsichord flatter if it is the upper note of a harmonic interval and the lower note is of a comparatively duller timbre. It is possible that participants were trying to compensate for the timbre difference that occurred in relation to the fixed reference note at C4. Since tuning one tone to another requires some form of perceptual matching to achieve a sound that is considered to be "in tune," it could be that participants tuned a brighter instrument flatter when comparing it to a duller instrument and tuned a duller instrument sharper when comparing it to a brighter instrument.

This experiment showed some findings that support the notion of timbre influencing tuning accuracy in harmonic intervals, especially at musically consonant intervals. Timbre pairs with greater distances between its instruments and presumably associated with decreased blend were more often significantly different from the 12TET standard at the unison, perfect fifth and octave. Although statistically significant comparisons were not found in support of tuning difficulty for any instrument combination, it is possible that the higher tuning variability in the 95% CI and standard deviations of dissonant intervals is a reflection of difficulty.

Chapter 3

Experiment 2: Timbre and Harmonic Interval Tuning 2

Based on the results of the first experiment, we loosely hypothesized that: 1) the upper (lower) note of a harmonic interval will be tuned flatter if brighter and sharper if duller than the lower (upper) note and 2) this effect, if it exists, would be stronger for musically consonant intervals. Additionally, it also investigated how different playing mechanisms affected tuning results, although we had no specific hypothesis for the effect of playing mechanism (impulsive vs. sustained) on tuning.

3.1 Method

3.1.1 Participants

Twenty musicians recruited from McGill University participated in this experiment (mean age = 23.1 years, SD = 4.9). All but one participant had over ten years of musical experience (mean years = 15.7, SD = 3.5) and at least one year of university-level ear training (mean years = 4.6, SD = 3.8). 85% of the musicians who participated in the study were multi-instrumentalists. In addition, 45% had conducting experience and 20% had composing or arranging experience. Out of all instruments mentioned, eight musicians had experience playing woodwinds (flute, bassoon, contrabassoon, clarinet, saxophone, oboe, recorder), sixteen had experience playing keyboards (piano), eight had experience playing percussion (drums, tabla, mallet percussion), seven had experience playing stringed instruments (guitars, cello, bass, viola, violin, ukulele), three had experience playing brass (trumpet, trombone, French horn, euphonium), one had experience playing electronic sounds and nine had experience performing with vocals.

After signing a written consent form, all participants met the required threshold of 20 dB HL on a pure-tone audiometric test with octave-spaced frequencies from 125 Hz to 8 kHz (ISO 389–8, 2004; Martin & Champlin, 2000) before beginning the first experimental session. After completing all of the experimental sessions, participants were debriefed and paid $30 CAD for roughly two to three hours of their time.

Twenty-one participants were run in total to obtain data from 20 analyzable subjects. One participant was dropped from the dataset after conducting a hierarchical cluster analysis based on a Pearson correlation matrix. A between groups average linkage dendrogram showed this person to be an outlier. Their average correlation with all other subjects was .03, whereas the average correlation of all other subjects excluding this participant was .1 (SD = .08). Although the outlying participant was slightly within 1 SD of the average correlation of all other subjects, their data were excluded based on the decision that there were too many extreme values (±50 cents) that did not correspond with the other subjects.

3.1.2 Stimuli

Twelve dyads composed of four different instruments were chosen based on brightness (bright vs. dull) and playing mechanism (impulsive vs. sustained). The four instruments were harpsichord (bright/impulsive), trumpet (bright/sustained), vibraphone (dull/impulsive) and flute (dull/sustained). The same flute sounds from Experiment 1 were used in Experiment 2, whereas the harpsichord, trumpet and vibraphone sounds were newly obtained from the Vienna Symphonic Library (https://vsl.co.at). The harpsichord for Experiment 2 was changed to a set of sounds recorded with single stops (one string per note), as the harpsichord from Experiment 1 had double stops (two strings per note). Like Experiment 1, the range of notes spanned an octave from C4 to C5 (specifically C4, C#4, E4, F#4, G4, A4, A#4 and C5), and the missing harpsichord notes for C4, E4, F#4, A#4 and C5 were transposed from the nearest existing semitone using Sound eXchange (http://sox.sourceforge.net). All pitches for each instrument were verified to match or closely match the fundamental frequencies of 12-TET with A4 at 440Hz using the same method from Experiment 1. Two pre-experiments were conducted. Five subjects adjusted the onset delays of the stimuli at each interval to increase their perceived onset synchrony. The median values across these subjects were used as the onset values for each instrument. Nine subjects participated in a loudness-matching experiment in which the perceived loudnesses of all stimuli were matched to the sound of Experiment 1's harpsichord at C4. Although Experiment 2's harpsichord sounds were different from Experiment 1, the harpsichord from Experiment 1 was used during loudness matching to retain the same loudness reference between the two experiments. Median loudness values calculated across all participants were used to adjust stimulus levels to achieve loudness equalization. The adjusted onset delay and loudness parameters were included in the tuning experiment.

3.1.3 Experimental Design

The experiment was a 12x8x2 repeated-measures design consisting of three within-subjects factors: 12 timbre pairs with different combinations of brightness and playing mechanism (TR/HC, HC/TR, HC/VB, VB/HC, FL/HC, HC/FL, VB/TR, TR/VB, TR/FL, FL/TR, VB/FL, FL/VB), eight pitch intervals (P1, m2, M3, A4, P5, m7, P8) and two tuning positions (low note, high note). The pitch intervals were chosen to have a mix of perfect consonant (P1, P5, P8), imperfect consonant (M3, M6) and dissonant (m2, A4, m7) intervals.

3.1.4 Procedure

Due to its overall duration, the experiment was split up into two sessions. Participants were alternately distributed into Group A (odd number) or B (even number), and were scheduled to come in twice over consecutive days or over three days to keep the time delay between sessions short. Group A, Session 1 consisted of 8 practice trials for each interval, 28 baseline trials for harpsichord and flute, and 96 experimental trials from pair order 1 tuned at both the upper and lower positions: TR/HC, HC/VB, FL/HC, VB/TR, TR/FL and VB/FL. Group A, Session 2 consisted of 28 baseline trials for trumpet and vibraphone, and 96 experimental trials from pair order 2: HC/TR, VB/HC, HC/FL, TR/VB, FL/TR and FL/VB. In both sessions, all baseline and experimental trials were randomly mixed into four blocks of 31 trials. Group B participants had the experimental sessions in inverted order other than the practice block at the start of Session 1. After completing the second session, all participants filled out an online demographic questionnaire. The interface and instructions mostly replicated that of Experiment 1 (see Appendix B, Experiment 1 Instructions, p. 78), with the only difference being that participants were also required to tune the lower note in Experiment 2. If the trial asked for an upper note adjustment, the lower note of the dyad was fixed at C4; if the trial asked for a lower note adjustment, the upper note was fixed at the note appropriate for that musical interval.

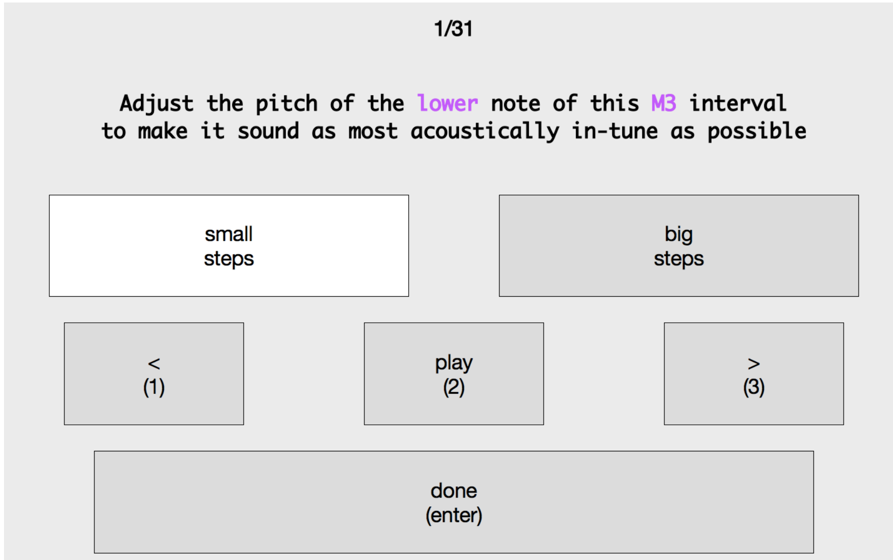

Fig. 3.1: A screenshot of the interface from Experiment 2.

3.2 Results

All analyses were measured against one dependent variable, Cent_Deviation. The first set of baseline analyses involved three independent variables (Instrument, Interval, Position), and the second set involved two (Instrument, Interval). The first set of experimental analyses involved three independent variables (Pair, Interval, Position), and the second set involved two (Pair, Interval). The data were analyzed with two different models using a GLM for Repeated Measures ANOVA in SPSS: 1) a 4 Instrument x 7 Interval x 2 Position baseline analysis along with its corresponding 12 Pair x 8 Interval x 2 Position experimental analysis and 2) a 4 Instrument x 7 Interval baseline analysis along with its corresponding 12 Pair x 8 Interval experimental analysis. Several series of paired-samples t-tests were also conducted to compare baseline and experimental means at the Bonferroni-adjusted alpha-level of $\alpha = .05/168 = .0003$ for each model.

3.2.1 Analysis of Baseline Means Including Position

Sphericity was violated for Instrument, $\chi^2(5) = 11.5$, $p = .04$, for Instrument*Interval, $\chi^2(170) = 290.7$, $p < .001$, for Instrument*Position, $\chi^2(5) = 18.6$, $p = .002$, for Interval*Position, $\chi^2(20) = 50.9$, $p < .001$ and for Instrument*Interval*Position, $\chi^2(170) = 364.7$, $p < .001$.

Greenhouse-Geisser corrections of ε = .71, ε = .36, ε = .66, ε = .55 and ε = .43 were used for each, respectively. Only the main effect of Interval was significant, $F(6, 114) = 2.6$, $p = .02$. Interaction effects between Instrument*Position, $F(1.98, 37.72) = 7.2$, $p < .001$ and Interval*Position, $F(3.31, 63.03) = 8.5$, $p < .001$, were significant, suggesting that in same-instrument conditions, tuning differences based on low or high note positions largely depended on either the instrument regardless of interval or the interval regardless of instrument. The three-way interaction between Instrument*Interval*Position was not significant. Bonferroni-adjusted pairwise comparisons between Instrument and Position showed that for HC/HC, the lower note was tuned –6.1 cents flatter than the upper note when averaged across all intervals, $p < .001$. For VB/VB, the lower note was tuned –3.1 cents flatter than the upper note, $p = .03$. As seen in Fig. 1, participants expanded HC/HC by flattening the lower note compared to a fixed upper note, whereas for VB/VB, participants expanded the interval by sharpening the upper note compared to a fixed lower note.

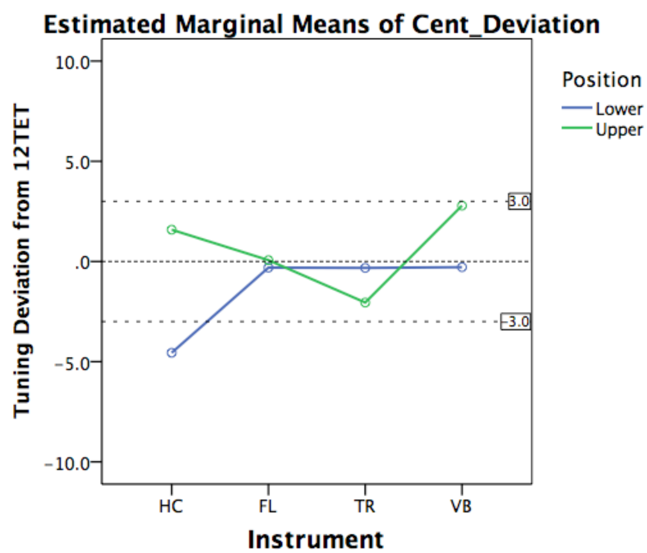

Fig. 3.2: Baseline mean cent deviations for Position as a function of Instrument.

Although the Instrument*Interval*Position interaction was not statistically significant overall, the profile plots of the mean cent deviations for Position from 12TET (0) as a function of Interval for each instrument tuned to itself displayed some visually interesting results.

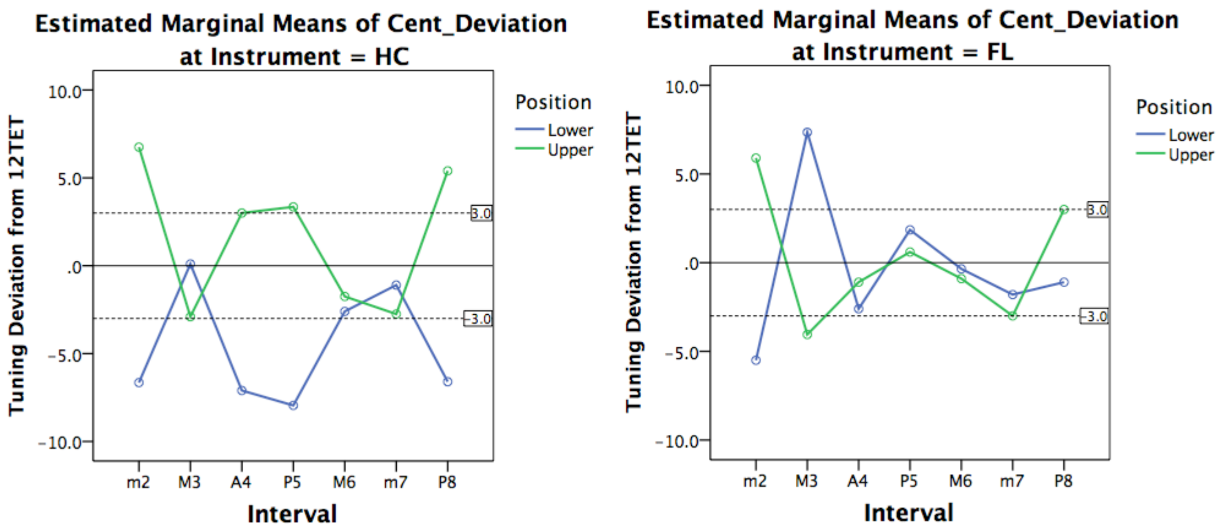

Fig. 3.3: Baseline mean cent deviations for Position as a function of Interval for HC and FL.

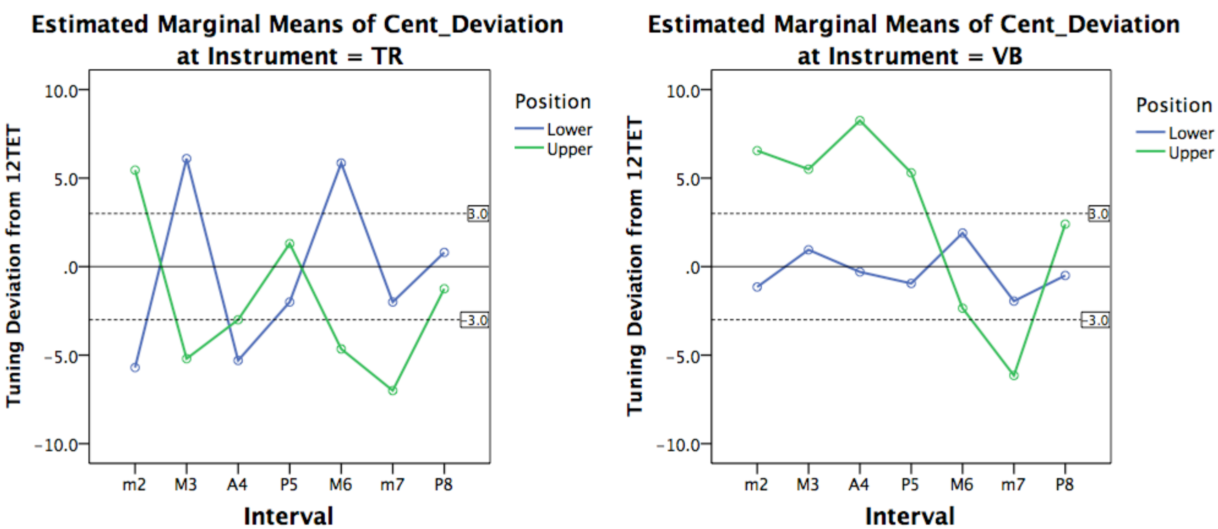

Fig. 3.4: Baseline mean cent deviations for Position as a function of Interval for TR and VB.

3.2.2 Analysis of Experimental Means Including Position

Sphericity values were unobtainable for Pair*Interval and Pair*Interval*Position, so Greenhouse-Geisser corrections of $\varepsilon = .16$ and $\varepsilon = .15$ were used for each, respectively. Only the main effect of Pair was significant, $F(11, 209) = 2.2$, $p = .01$. The interaction effects of Pair*Position, $F(11, 209) = 14.8$, $p < .001$, Interval*Position, $F(7, 133) = 5.8$, $p < .001$ and Pair*Interval*Position, $F(11.95, 227.02) = 2.1$, $p = .02$, were significant, which were further investigated using Bonferroni-adjusted pairwise comparisons.

The interaction effect between Pair and Position showed that when averaged across all intervals, participants either expanded or contracted the interval by flattening or sharpening the lower instrument, respectively. Participants tuned the trumpet of TR/HC sharper than the harpsichord of HC/FL. For the HC/TR pairing in particular, the effect of interval expansion by flattening the lower note was greater in comparison to other instrument pairings. Some of the more interesting differences include the flattening of the harpsichord of HC/TR compared to the trumpet of TR/HC, the vibraphone of VB/HC and the flute of FL/HC and FL/TR. Participants also tuned the harpsichord of HC/VB flatter than the flute of FL/HC. Finally, in a similar manner to the results found in Experiment 1, there was a significant difference in instrument order for the pairs FL/HC and HC/FL. Participants tuned the flute of FL/HC sharper than the harpsichord of HC/FL, contracting the interval when tuning the lower note of FL/HC in comparison to when they were tuning the lower note of HC/FL. The four pairs that showed order importance when tuning the lower note were TR/HC, HC/TR, FL/HC and HC/FL.

Averaged across all intervals, participants also either expanded or contracted the interval by sharpening or flattening the upper instrument, respectively. Participants contracted the interval when tuning the harpsichord of TR/HC in comparison to all other pairings, notably tuning it flatter than the trumpet of HC/TR, the vibraphone of HC/VB and the flute of HC/FL and TR/FL. In contrast, participants expanded the interval when tuning the trumpet of HC/TR in comparison to all other pairings, notably tuning it sharper than the harpsichord of VB/HC and FL/HC, the vibraphone of TR/VB and the trumpet of FL/TR. Interval contraction occurred with the flattening of the harpsichord for FL/HC compared to the flute of HC/FL and VB/FL, whereas interval expansion occurred with the sharpening of the flute for HC/FL compared to the trumpet of FL/TR. Participants also tuned the vibraphone of TR/VB flatter than the flute of VB/FL and tuned the trumpet of FL/TR flatter than the flute of VB/FL. Interestingly, order difference was seen for the pairs VB/FL and FL/VB, where participants tuned the flute of VB/FL sharper than the vibraphone of FL/VB. The pairs that showed order importance when tuning the upper note were TR/HC, HC/TR, FL/HC, HC/FL, VB/FL and FL/VB.

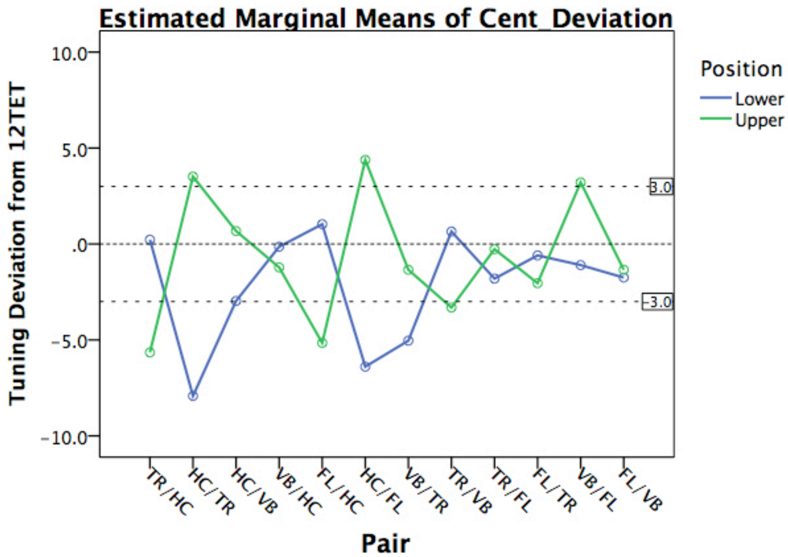

Fig. 3.5: Experimental mean cent deviations for Position as a function of Pair.

Some pairs showed significantly different tuning directions between their lower and upper note positions when averaged across all intervals. Opposite tuning directions suggest either the same contraction (sharpen low note and flatten upper note) or expansion (flatten low note and sharpen upper note) effect for each specific pair, and these intra-pair effects corresponded to the inter-pair effects seen in comparisons between pairs. Interval contraction occurred for TR/HC, TR/VB and FL/HC, whereas interval expansion occurred for HC/TR, HC/VB and HC/FL.

A closer look at the three-way interaction between Pair, Interval and Position revealed that significant differences between pairs for both lower and upper note tuning positions occurred only at P1, M3, P5 and P8. Significant interval contractions occurred for pairs TR/HC, VB/HC and FL/HC, and significant interval expansions occurred for pairs HC/TR, HC/VB and HC/FL. For lower note tunings at P1, participants contracted TR/HC by a greater amount compared to many other pairings, but there were no significant differences between the trumpet of TR/HC and the vibraphone and flute of VB/HC and FL/HC, respectively. Participants contracted VB/HC compared to HC/TR, HC/VB, HC/FL and VB/TR; and FL/HC compared to HC/TR, HC/VB, HC/FL, TR/FL, FL/TR and VB/FL. Participants expanded HC/TR compared to TR/HC, VB/HC, FL/HC, TR/VB, TR/FL and FL/TR; HC/VB compared to TR/HC, VB/HC, FL/HC and TR/VB; and HC/FL compared to TR/HC, VB/HC, FL/HC and TR/FL. For upper note tunings at P1, participants contracted TR/HC in comparison to all other pairings; VB/HC compared to HC/TR, HC/VB, HC/FL and VB/FL; and FL/HC compared to HC/TR, HC/VB, HC/FL and VB/FL. Participants expanded HC/TR compared to TR/HC, VB/HC, FL/HC, TR/VB and TR/FL; HC/VB compared to TR/HC, VB/HC, FL/HC, TR/VB and FL/VB; and HC/FL compared to TR/HC, VB/HC, FL/HC, TR/FL, FL/TR and FL/VB.

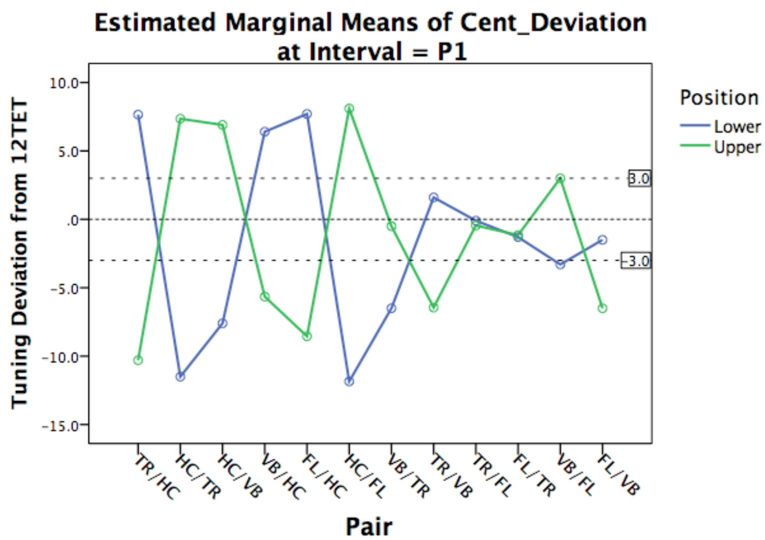

Fig. 3.6: Experimental mean cent deviations for Position as a function of Pair at P1.

There were no lower note tuning differences seen at M3. For upper note tunings, participants contracted FL/HC compared to HC/FL, and expanded HC/FL compared to FL/HC and FL/TR.

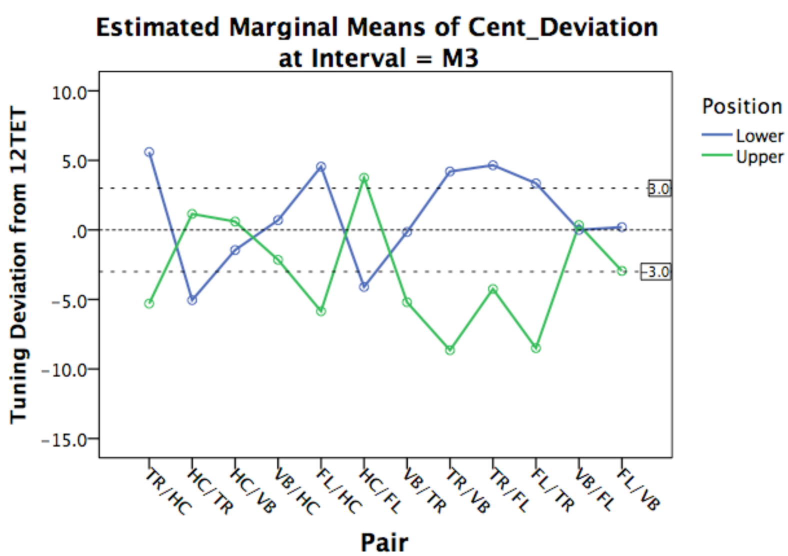

Fig. 3.7: Experimental mean cent deviations for Position as a function of Pair at M3.

For lower note tunings at P5, participants contracted TR/HC compared to HC/TR. In contrast, participants expanded HC/TR compared to TR/HC, HC/VB, VB/HC, FL/HC, TR/VB, TR/FL and FL/TR. For upper note tunings at P5, participants contracted TR/HC compared to HC/TR and HC/FL, and expanded HC/TR compared to TR/HC, FL/HC and FL/TR. Participants also contracted FL/HC compared to HC/TR, HC/FL and VB/FL.

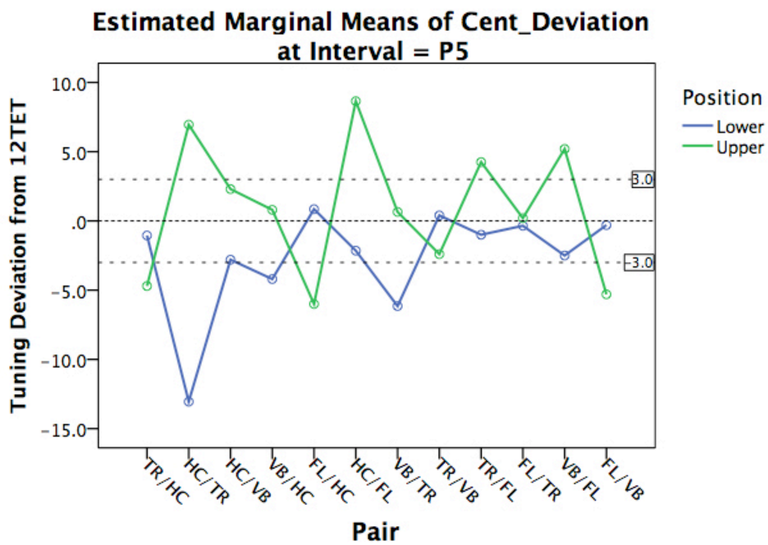

Fig. 3.8: Experimental mean cent deviations for Position as a function of Pair at P5.

For lower note tunings at P8, participants contracted TR/HC compared to HC/TR and HC/FL, and expanded HC/TR compared to TR/HC, VB/HC, FL/HC, TR/VB, TR/FL and FL/TR. Participants expanded HC/VB compared to TR/VB, and contracted VB/HC compared to HC/TR and HC/FL. Participants contracted FL/HC compared to HC/TR and HC/FL, and expanded HC/FL compared to TR/HC, VB/HC, FL/HC, FL/TR, FL/VB, TR/FL and VB/FL. For upper note tunings at P8, participants contracted TR/HC compared to HC/TR and HC/FL, and expanded HC/TR compared to TR/HC and TR/VB. Participants expanded HC/VB by tuning the vibraphone sharper than the harpsichord of FL/HC, and contracted VB/HC by tuning the harpsichord flatter than the flute of HC/FL. Lastly, participants contracted FL/HC compared to HC/VB and HC/FL, and expanded HC/FL compared to TR/HC, VB/HC and FL/HC.

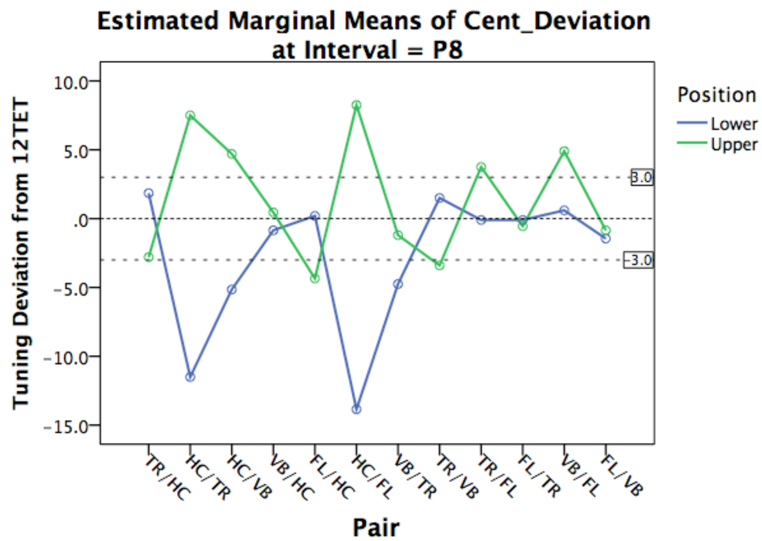

Fig. 3.9: Experimental mean cent deviations for Position as a function of Pair at P8.

There were notable upper and lower tuning position differences at certain intervals for some pairs. For TR/HC, participants contracted the interval by sharpening the trumpet compared to the harpsichord at P1 and M3. HC/TR showed clear instances of interval expansion at every interval except m2 and m7, where participants flattened the harpsichord compared to the trumpet. HC/VB showed interval expansion at intervals P1, P5 and P8, where participants flattened the harpsichord compared to the vibraphone. For VB/HC, participants contracted the interval by tuning the vibraphone sharper than the harpsichord at P1 and m7, but expanded the interval at M6 by tuning the vibraphone flatter than the harpsichord. Participants contracted the interval for FL/HC at P1, M3 and P5 by tuning the flute sharper than the harpsichord. Participants expanded the interval for HC/FL at P1, M3, P5, m7 and P8, tuning the harpsichord flatter than the flute. For TR/VB, participants tuned the trumpet sharper than the vibraphone at P1 and M3. For TR/FL, participants tuned the trumpet sharper than the flute at M3, but flatter at P8. For FL/TR, participants tuned the flute sharper than the trumpet at M3. For VB/FL, participants tuned the vibraphone flatter than the flute at P1 and A4.

3.2.3 Comparison of Baseline and Experimental Means Including Position

Paired-samples t-tests were conducted to compare baseline and experimental means. Only the pairs with the same instrument tuned at either the upper or lower note positions were analyzed. For example, lower note tunings of HC/HC were compared to pairs HC/TR, HC/VB and HC/FL, and upper note tunings of HC/HC were compared to pairs TR/HC, VB/HC and FL/HC. Because baseline measures were not recorded for interval P1, intervals m2 through P8 were compared. Only the p-values at $\alpha = .0003$ and below were declared significant.

For harpsichord pairs, only the upper note tuning comparisons were significant. At m2 and P5, the upper note of HC/HC was tuned sharper than the upper note of TR/HC, whereas at P5 and P8, it was tuned sharper than the upper note of FL/HC.

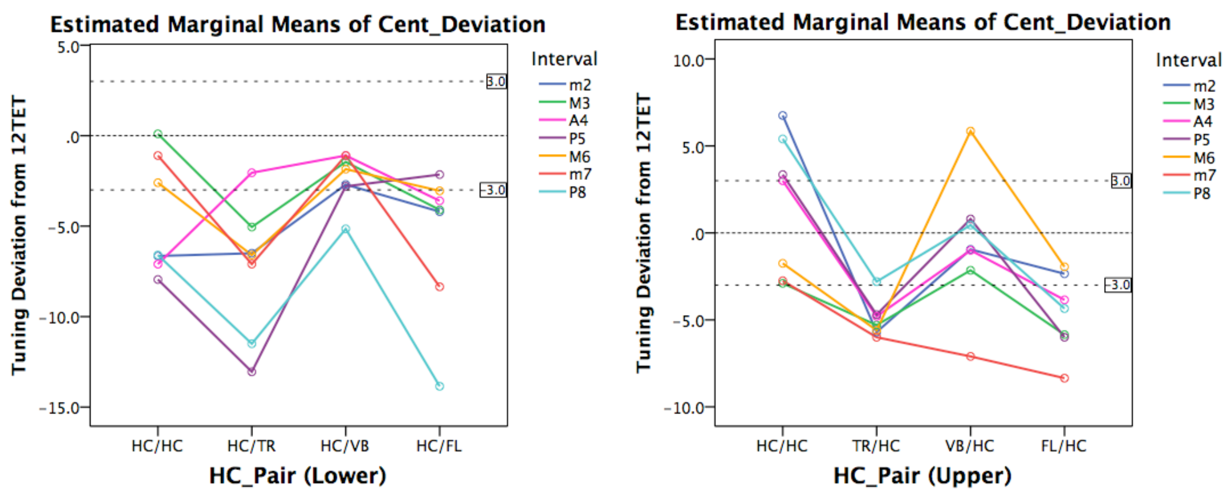

Fig. 3.10: Baseline and experimental mean cent deviations for Harpsichord.

There was only one significant difference for the set of trumpet pairs tuned at the upper note. At P8, the upper note of TR/TR was tuned flatter than the upper note of HC/TR.

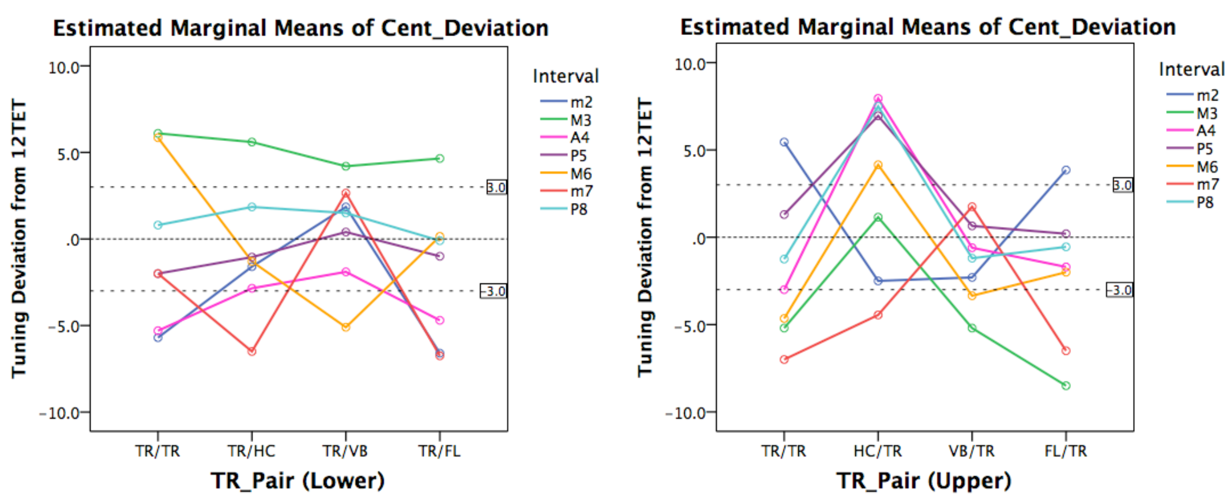

Fig. 3.11: Baseline and experimental mean cent deviations from 12TET (0) for Trumpet.

There were no significant mean differences for the set of flute and vibraphone pairs tuned at either the lower or upper note under the Bonferroni adjustment.

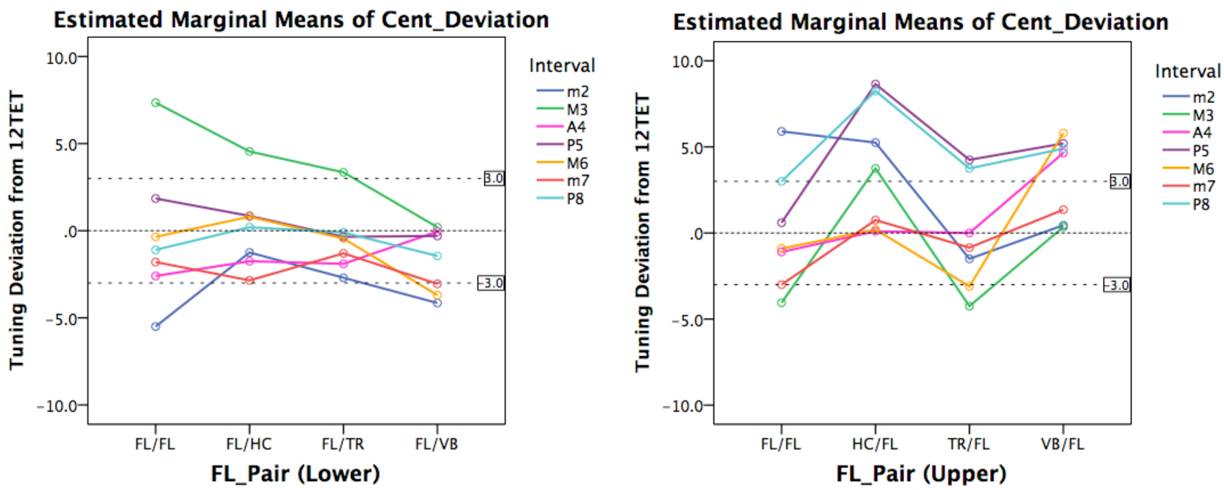

Fig. 3.12: Baseline and experimental mean cent deviations for Flute.

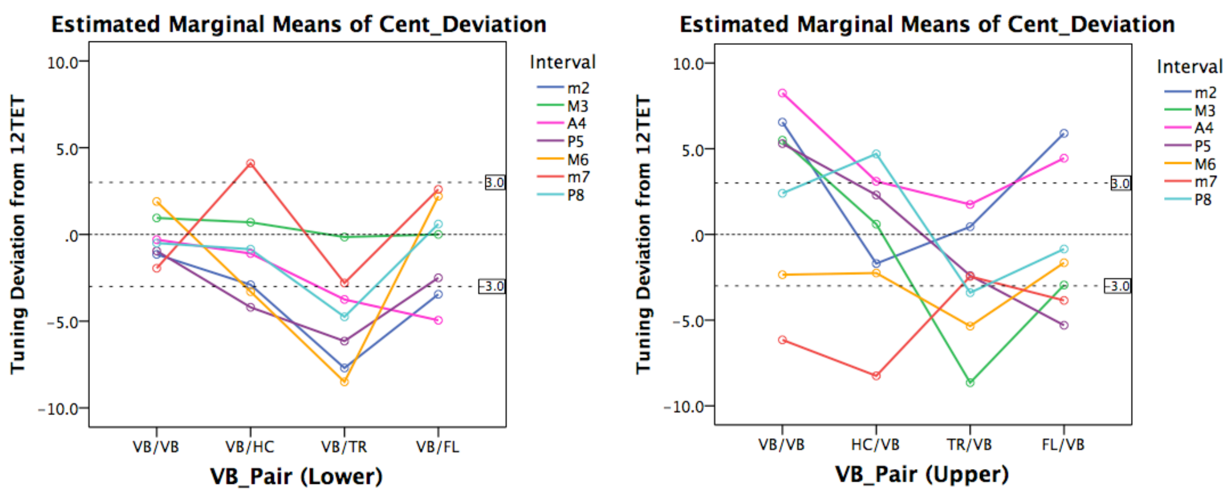

Fig. 3.13: Baseline and experimental mean cent deviations for Vibraphone.

3.2.4 Analysis of Baseline Means Excluding Position

To see if removing the Position factor would improve the clarity of results, a second set of repeated measures ANOVA was run on the data. The Position factor was removed by inverting the value of the measures obtained for low note tunings to reflect expansion / contraction of the interval size (e.g., if participants sharpened the low note, which would be recorded as a positive value, it symbolizes a contraction of the interval, which can be represented by a negative value). The means across both high and low note tunings were then calculated for each subject to get a single value reflecting total change in interval size from equal temperament (+0) regardless of tuning position.

Sphericity was violated for Instrument, $\chi^2(5) = 18.5$, $p = .002$, Interval, $\chi^2(20) = 50.9$, $p < .001$ and Instrument*Interval, $\chi^2(170) = 364.7$, $p < .001$. Greenhouse-Geisser values of $\varepsilon = .66$, $\varepsilon = .55$ and $\varepsilon = .43$ were considered for each, respectively. The main effects of Instrument, $F(1.98, 37.72) = 7.19$, $p = .002$ and Interval $F(3.32, 63.03) = 8.52$, $p < .001$, were significant. There was no interaction effect between Instrument and Interval. The main effect of Instrument showed that all of the significant differences were in relation to the HC/HC pair. Participants expanded the interval size for HC/HC by 2.9 cents greater than FL/FL, $p < .001$ and 3.9 cents greater than TR/TR, $p < .001$. There were no differences found between VB/VB and any other instrument pair. As found in the previous 4 Instrument x 7 Interval x 2 Position baseline analysis, there was no interaction effect found between Instrument and Interval. Since most of the previous significant differences came about in relation to the factor Position, removing it to reflect a total change in interval size only affected the main effect of Instrument. As seen in Fig. 2.13, the four baseline pairs follow a similar contour across intervals, which is especially so for the two bright instruments, harpsichord and trumpet.

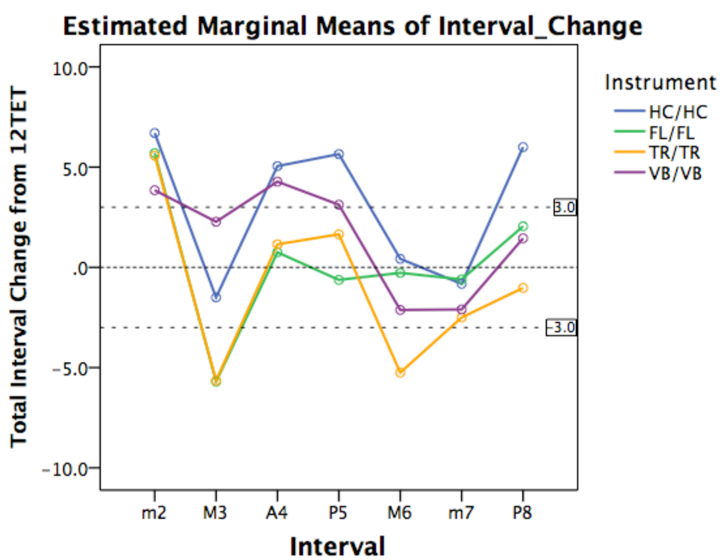

Fig. 3.14: Baseline mean change in interval size for Instrument as a function of Interval.

3.2.5 Analysis of Experimental Means Excluding Position

A sphericity value was unobtainable for Pair*Interval, so the Greenhouse-Geisser correction of $\varepsilon = .15$ was used. Both main effects of Pair, $F(11, 209) = 14.8$, $p < .001$ and Interval, $F(7, 133) = 5.8$, $p < .001$, as well as the interaction effect of Pair*Interval, $F(11.95, 227.02) = 2.1$, $p = .02$ were found to be significant. As found in the 12 Pair x 8 Interval x 2 Position experimental measures analysis, all of the significant differences for the main effect of pair across all intervals were found in relation to the six pairs containing the harpsichord: TR/HC, HC/TR, HC/VB, VB/HC, FL/HC and HC/FL. Most of these differences stemmed from the first two pairs, TR/HC and HC/TR. Of note, TR/HC showed interval contraction compared to HC/TR, HC/VB, HC/FL, VB/TR and TR/FL. HC/TR showed significant interval expansion compared to TR/HC, HC/VB, VB/HC, FL/HC, TR/VB, TR/FL and FL/TR. Pair HC/VB showed interval expansion compared to TR/HC, FL/HC and TR/VB, but it showed interval contraction compared to HC/TR. Pair VB/HC showed interval contraction compared to HC/TR and HC/FL. Pair FL/HC showed interval contraction compared to HC/TR, HC/VB, HC/FL, TR/FL, VB/FL and FL/VB. Pair HC/FL showed interval expansion compared to TR/HC, VB/HC, FL/HC, TR/FL and FL/TR.

The interaction effect between Pair and Interval showed that, as expected, all of the significant differences were found in relation to the six pairs including the harpsichord at intervals P1, M3, P5 and P8. In general, results suggested that having the harpsichord as the upper or lower note affects tuning results similarly regardless of the instrument playing the lower or upper note, respectively. To a lesser extent, this also seemed to be the case for pairs that had a trumpet as the lower or upper note with a percussive instrument playing the upper or lower note, respectively. The only pairs that TR/HC was not significantly different from at P1 were VB/HC, FL/HC, TR/VB and FL/VB. TR/HC showed interval contraction compared to HC/TR, HC/VB, HC/FL, VB/TR, TR/FL and FL/TR. HC/TR was significantly different from every other pair at P1 except HC/VB, HC/FL and VB/TR. HC/TR showed interval expansion compared to TR/HC, VB/HC, FL/HC, TR/VB, TR/FL and FL/TR. HC/VB was significantly different from every other pair at P1 except HC/TR, HC/FL, VB/TR and VB/FL. In particular, HC/VB showed interval expansion compared to TR/HC, VB/HC, FL/HC, TR/VB and FL/VB. VB/HC showed interval contraction compared to HC/TR, HC/VB, HC/FL and VB/FL. FL/HC was significantly different from every other pair at P1 except TR/HC, VB/HC, TR/VB and FL/VB. It is possible that the lack of difference between FL/HC and FL/VB stems from the fact that the instruments playing the upper note were both percussive. FL/HC showed interval contraction compared to HC/TR, HC/VB, HC/FL, TR/FL, FL/TR and VB/FL. HC/FL was significantly different from every other pair at P1 except HC/TR, HC/VB, VB/TR and VB/FL. HC/FL showed interval expansion compared to TR/HC, VB/HC, FL/HC, TR/FL and FL/TR. TR/HC was only significantly different from HC/TR at M3, where it showed interval contraction compared to HC/TR. HC/TR showed interval expansion compared to TR/HC, TR/VB, TR/FL and FL/TR. Although there were no significant differences between FL/HC with other pairs at M3, HC/FL showed interval expansion compared to TR/FL and FL/TR.

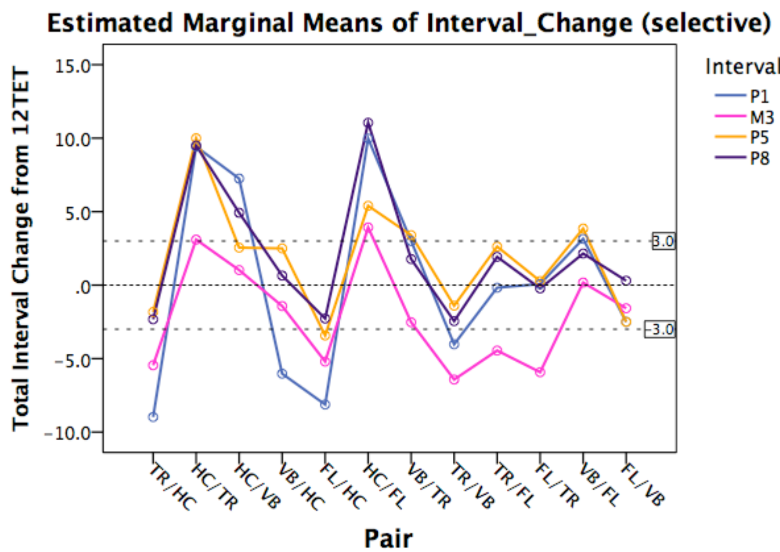

Fig. 3.15: Experimental mean change in interval size for intervals P1, M3, P5 and P8 as a function of Pair.

At P5, TR/HC showed interval contraction compared to HC/TR. HC/TR was significantly different from every other pair except HC/FL, VB/TR, TR/FL and VB/FL. HC/TR showed interval expansion compared to TR/HC, HC/VB, VB/HC, FL/HC, TR/VB and FL/TR. Both HC/VB and VB/HC showed interval contraction compared to HC/TR. FL/HC showed interval contraction compared to HC/TR and HC/FL. HC/FL showed interval expansion compared to FL/HC. At P8, TR/HC showed interval contraction compared to HC/TR and HC/FL. HC/TR was significantly different from every other pair except HC/VB and HC/FL, showing interval expansion compared to TR/HC, VB/HC, FL/HC, VB/TR, TR/VB, TR/FL and

FL/TR. HC/VB showed interval expansion compared to FL/HC, and VB/HC showed interval contraction compared to HC/TR and HC/FL. FL/HC showed interval contraction compared to HC/TR, HC/VB and HC/FL. HC/FL was significantly different from every other pair except for HC/TR and HC/VB. HC/FL showed interval expansion compared to TR/HC, VB/HC, FL/HC, TR/FL, FL/TR, VB/FL and FL/VB.

3.2.6 Comparison of Baseline and Experimental Means Excluding Position

Another series of Bonferroni-corrected paired-samples t-tests were conducted to compare baseline and experimental means for intervals m2 through P8. Because Position was excluded as a factor, each baseline instrument pair was compared with all other pairs that contained the same instrument in either the lower or upper note position. Only the p-values at $\alpha = .0003$ or below were declared significant.

HC/HC was significantly different from TR/HC and FL/HC at P5 and P8, where HC/HC showed interval expansion compared to TR/HC and FL/HC.

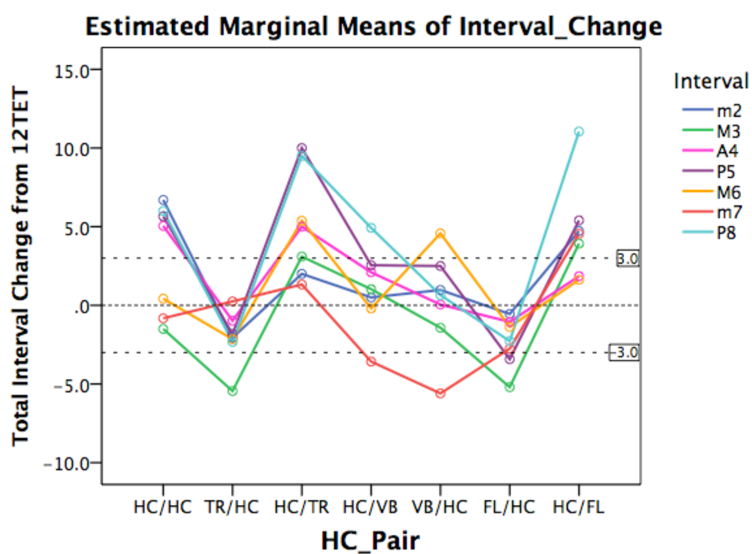

Fig. 3.16: Baseline and experimental mean change in interval size for Harpsichord.

FL/FL was significantly different from only HC/FL at M3 and P8, where FL/FL showed interval contraction compared to HC/FL.

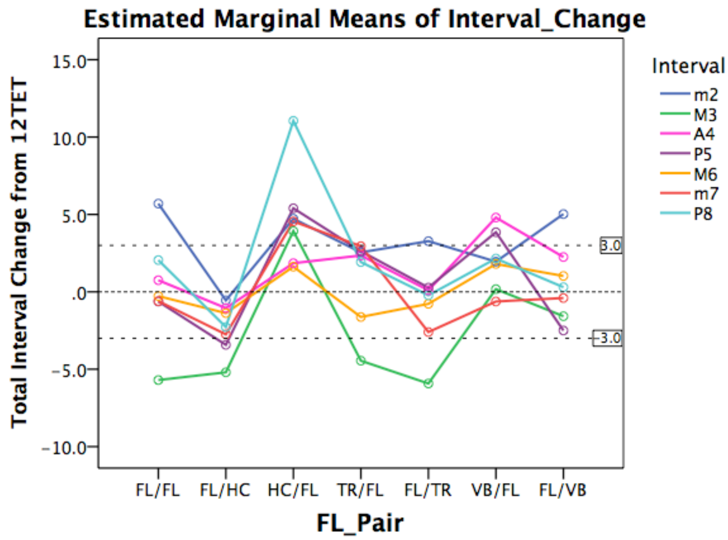

Fig. 3.17: Baseline and experimental mean change in interval size for Flute.

TR/TR was significantly different from only HC/TR at M3, P5 and P8, where TR/TR showed interval contraction compared to HC/TR.

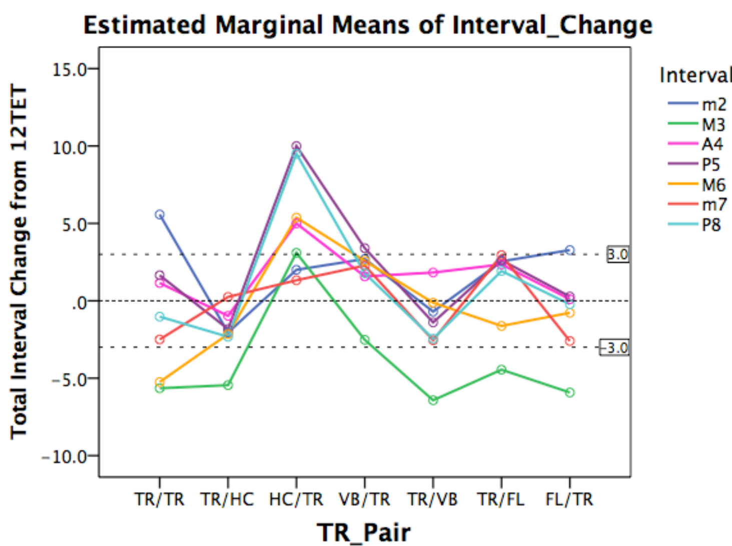

Fig. 3.18: Baseline and experimental mean change in interval size for Trumpet.

VB/VB was not significantly different from any other vibraphone pair at any interval, which was also the case for the comparison between baseline and experimental means including the Position factor.

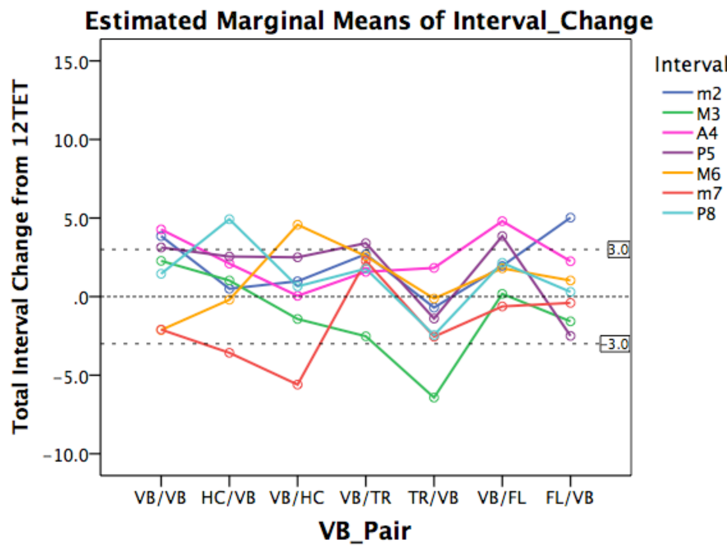

Fig. 3.19: Baseline and experimental mean change in interval size for Vibraphone.

3.3 Discussion

Results showed that, rather than an effect of brightness or dullness on tuning harmonic intervals, there were specific effects related to the harpsichord that occurred primarily around musically consonant intervals. The harpsichord, which is a bright, impulsive instrument, prompted musicians to contract the interval at P1, M3, P5 and P8 if it was playing the upper note, and expand the interval at P1, M3, P5 and P8 if it was playing the lower note regardless of tuning position (the note being tuned) and the instrument playing the other note. When compared to all other pairs at P1, interval contraction occurred for TR/HC, VB/HC and FL/HC, and interval expansion occurred for HC/TR, HC/VB and HC/FL. At M3, interval contraction occurred only for FL/HC, and interval expansion occurred only for HC/FL. At P5, interval contraction occurred for TR/HC and FL/HC, and interval expansion occurred for HC/TR and HC/FL. At P8, interval contraction occurred for TR/HC, VB/HC and FL/HC, and interval expansion occurred for HC/TR, HC/VB and HC/FL.

Regardless of tuning position, pairs TR/HC, VB/HC and FL/HC were significantly different from other pairs but never significantly different from each other at any interval. Pairs HC/TR, HC/VB and HC/FL were also not significantly different from each other at any interval for the analyses that included the Position factor, but HC/TR and HC/VB were significantly different from each other at interval P5 for the analyses that excluded the Position factor. Despite this, the results give good implications that the harpsichord was the strongest catalyst for interval contraction or expansion by affecting tuning results similarly between pairs that had the harpsichord in the same note position (low or high). Instrument order effects were prevalent for these six harpsichord pairs, showing significantly different results between pairs that had the harpsichord in opposite note positions (low vs. high). These results specific to the harpsichord are markedly different from previous research findings demonstrating that musicians produce sharper intonation when performing out of tune (Geringer & Witt, 1985), especially when playing the lower note of a harmonic interval (Karrick, 1998). It must be noted, though, that these studies were based more on intonation during performance, and looked at a different set of instrumental stimuli.

Technically, given that the harpsichord stimuli used in this experiment were comparatively brighter than the other stimuli, the hypotheses made a priori about the effects of brightness on tuning were not completely disproved. However, they are only applicable to the pairs that contain the harpsichord. Significant tuning position differences within pairs indicate opposite tuning directions for the lower and upper notes, which shows either the same contraction (sharpen lower note and flatten upper note) or expansion (flatten lower note and sharpen upper note) effect for each pair. Position differences within pairs showed that the six harpsichord pairs were in agreement with the effects found from comparisons between pairs, where having the harpsichord as the upper note prompted an interval contraction and having it as the lower note prompted an interval expansion. If this were merely due to brightness, other pairs that showed position differences within themselves would have followed the same pattern, but they did not. TR/VB showed interval contraction at P1 and M3 despite having a duller instrument as the upper note. TR/FL showed interval contraction at M3 and interval expansion at P8. FL/TR showed interval contraction at M3, and VB/FL showed interval expansion at P1 and A4. Perhaps the harpsichord stimuli had something to do with increased sensory dissonance based on the incongruity between perceived and expected templates of common chords (McLachlan et al., 2013) and reduced blend due to its relatively inharmonic nature (Moore et al., 1986) in comparison to the other instruments.

For the comparison analyses between baseline and experimental measures that included Position, the only baseline pairs that were tuned significantly differently from their experimental counterparts were HC/HC and TR/TR, and the differences arose only for upper note tuning means. For the set of comparison analyses that excluded Position, HC/HC, FL/FL and TR/TR were also only significantly different in interval contraction/expansion from pairs that had the same instrument in the upper note position (i.e., TR/HC, FL/HC, HC/FL and HC/TR, respectively). Despite the fact that many participants reported difficulty tuning vibraphone pairs whether through instrument unfamiliarity or aversion, vibraphone pairs displayed some of the most consistently accurate means in relation to 12TET (0). However, vibraphone pairs also displayed the widest confidence interval bounds compared to the other instruments, suggesting a larger range of tuning variability. As concluded in Experiment 1, tuning difficulty is not necessarily correlated with tuning inaccuracy from equal temperament, but seems to be associated with tuning variability.

Based on the results, the effects of brightness and playing mechanism on harmonic interval tuning are inconclusive. However, the effects of the harpsichord alone were surprisingly (or perhaps, unsurprisingly) consistent at musically consonant intervals. It is unclear if this is due to its brightness, its inharmonicity, its playing mechanism, or an amalgamation of these properties. What is interesting to note is that although the timbre of the harpsichord samples used in Experiment 2 were different from those used in Example 1, they produced similar tuning results in similar circumstances.

Chapter 4

Experiment 3: Timbre and Harmonic Interval Perception

The primary goal of Experiment 3 was to test whether or not certain instrument pairs from the entire set of stimuli from Experiment 1 and 2 are considered to be more or less blended than others even at the same musical intervals due to their varying degrees of brightness and playing mechanism. Based on Plomp and Levelt (1965), additional ratings on auditory roughness and pleasantness were included in the design for further investigation into the perceptual interaction between sensory dissonance and musical consonance. Since pairs FL/HC and HC/FL were specifically chosen based on their comparatively far distance in the Lakatos (2000) combined timbre space and based on Kendall and Carterette's (1993) finding that increased distance in a timbre space decreases blend, we hypothesized that participant ratings would reflect these two pairs as the least blended compared to other pairs. Because Sandell (1995) found that increased composite and average brightness as well as decreased onset synchrony reduces blend, we also hypothesized that pairs with at least two of the following properties of opposite playing mechanisms (impulsive vs. sustained), high total brightness (bright + bright) and large differences in brightness (bright vs. dull) would be comparatively less blended than other pairs. Such pairs under these criteria include FL/HC, HC/FL, TR/HC, HC/TR, TR/VB, VB/TR, TR/PN and PN/TR. It was also postulated that pairs perceived to be less blended in this study would be correlated with the pairs that had greater tuning variability in the previous experiments.

4.1 Method

4.1.1 Participants

Twenty musicians recruited through McGill University participated in this experiment (mean age = 24.5, SD = 3.9). The average amount of musical experience across all participants was 17.2 years (SD = 4.3) with 3.6 years of university-level ear training (SD = 1.8). 90% of the musicians who participated in the study were multi-instrumentalists. In addition, 25% had conducting experience and 20% had composing or arranging experience. Out of all instruments mentioned, eight musicians had experience playing woodwinds (flute, bassoon, contrabassoon, saxophone, clarinet), eight had experience playing stringed instruments (guitars, violin, viol, viola, cello, bass), five had experience playing percussion (drums, tabla), sixteen had experience playing keyboards (piano, harpsichord), five had experience playing brass (French horn, trumpet, trombone, tuba), one had experience playing electronic sounds and seven had experience performing with vocals. After signing a written consent form, all participants met the required threshold of 20 dB HL on a pure-tone audiometric test with octave-spaced frequencies from 125 Hz to 8 kHz (ISO 389–8, 2004; Martin & Champlin, 2000) before beginning the experiment. After completing the session, participants were debriefed and paid $10 CAD for roughly 30 to 50 minutes of their time.

Twenty-one participants were run in total to obtain data from 20 analyzable subjects. One participant was dropped from the dataset after a hierarchical cluster analysis based on a Pearson correlation matrix and a between-groups average linkage dendrogram showed the participant to be an extreme outlier. Their average correlation with all other participants was .04, and the average correlation of all other subjects excluding the outlier was .38 (SD = .1). This participant was roughly 3.5 SDs below the average correlation of all other subjects.

4.1.2 Stimuli

All stimuli used in Experiment 3 were taken from the previous tuning experiments, with the harpsichord being the one used in Experiment 2. A series of one-sample t-tests against zero (equal temperament) were conducted in order to determine if it was necessary to adjust the tuning of each pair to reflect the average tuning results from Experiments 1 and 2. For Experiment 1, seven out of 76 differences were significant under the Bonferroni-adjusted alpha-level of .0006, where the mean difference range of these significant instances was 3.95 to 14.1 cents. For Experiment 2, 15 out of 248 differences were significant under the Bonferroni-adjusted alpha-level of .0002, where the mean difference range of these significant instances was 5.4 to 13.85 cents. For both experiments, even the pairs that were the most significantly different from equal temperament remained within the average just noticeable difference of 18 cents between the natural and tempered scales for harmonic intervals as found by Moran and Pratt (1926). It was thus decided that the stimuli used for Experiment 3 did not need to be altered from their original equal-tempered tuning.

4.1.3 Experimental Design

The experiment was a 16x8 repeated-measures design consisting of two within-subjects factors: 16 timbre pairs with different combinations of timbre space distance, brightness and playing mechanism (PN/HC, HC/PN, FL/CL, CL/FL, FL/HC, HC/FL, TR/HC, HC/TR, TR/FL, FL/TR, TR/VB, VB/TR, VB/HC, HC/VB, FL/VB, VB/FL) and eight pitch intervals (P1, m2, M3, A4, P5, m7, P8) measured on three different dependent variables: roughness, blend and pleasantness.

4.1.4 Procedure

The experiment was divided into five blocks. The first block consisted of eight practice trials and the remaining four consisted of 44 randomized baseline and experimental trials each, for a total of 184 trials. After completing the session, all participants filled out an online questionnaire for demographic purposes. The sound levels and experiment setup remained identical to the ones used in Experiments 1 and 2. The program interface was modified to include instructions at the top and three continuous scales beneath.

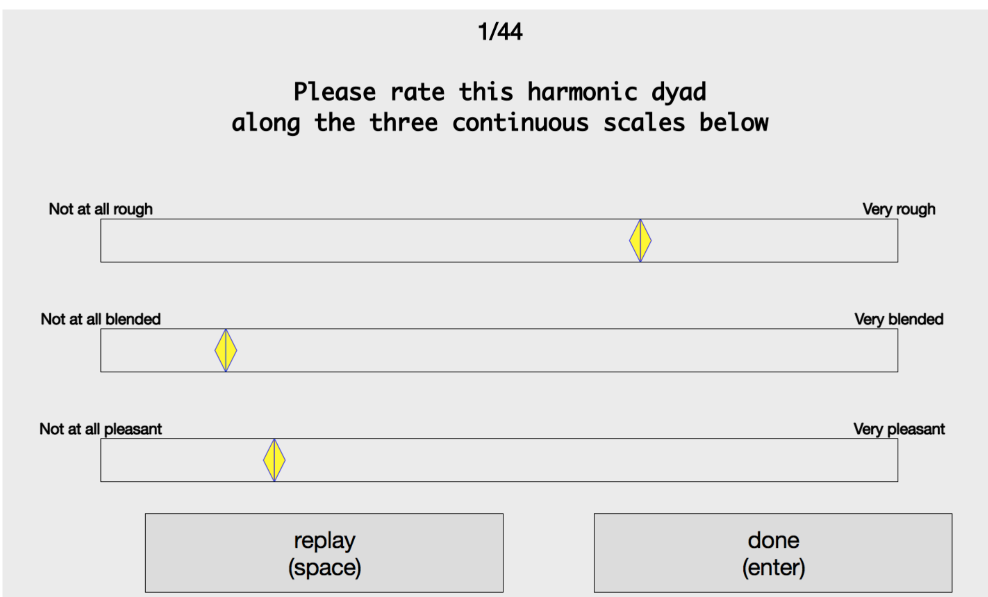

Fig. 4.1: A screenshot of the interface from Experiment 3.

Participants were familiarized with definitions and examples for the terms "auditory roughness" and "blend" before starting the experiment (see Appendix C, Experiment 3 Instructions, p. 79). They were instructed to make subjective ratings on the auditory roughness, blend and pleasantness of various harmonic dyads along three scales labeled, “Not at all rough” to “Very rough,” “Not at all blended” to “Very blended,” and “Not at all pleasant” to “Very pleasant.” They were also told to focus on the overall auditory characteristics of each sound event rather than just the instrumental timbres or the musical intervals. A ‘Replay’ button allowed participants to listen to the dyad a maximum of three additional times if necessary. Once participants were satisfied with their ratings, they were required to click ‘Done’ or press the ‘enter’ key to move on to the next trial.

4.2 Results

Baseline analyses involved two independent variables (Instrument, Interval) and three dependent variables (roughness, blend, pleasantness), whereas experimental analyses involved two independent variables (Pair, Interval) with the same three dependent variables. Dependent variables were coded along arbitrary scales of 0 (minimum) to 1 (maximum). Since roughness and blend were not normally distributed, their ratings were analyzed differently compared to pleasantness. For each set of baseline (6 Instrument x 8 Interval) and experimental (16 Pair x 8 Interval) measures, roughness and blend were analyzed together in a doubly multivariate repeated-measures design using the GLM Repeated Measures option in SPSS. The doubly multivariate repeated-measures model was used because each subject made three separate scale ratings (two of which may be related) for every trial (factorial combination of all pairs and intervals). Subsequent analyses were made using separate univariate repeated measures models for each dependent variable to investigate the Bonferroni-corrected pairwise comparisons of the significant multivariate effects. A doubly multivariate repeated-measures model was also used to compare differences between baseline and experimental means for roughness and blend, whereas a univariate approach was taken to compare the differences between baseline and experimental means for pleasantness. Finally, bivariate Pearson correlations were conducted between the dependent variables of Experiment 1 and 2 with the corresponding perceptual ratings from Experiment 3.

4.2.1 Analysis of Baseline Means for Roughness and Blend

Sphericity values were ignored because multivariate analyses do not require sphericity to be met. There was a significant multivariate effect of roughness and blend as a group on Instrument, Wilk's Lambda (λ) = .37, $F(10, 188) = 11.8$, $p < .001$ and on Interval, Wilk's λ = .23, $F(14, 264) = 20.1$, $p < .001$. There was also a significant multivariate effect on the interaction effect of Instrument*Interval, Wilk's $\lambda = .87$, $F(70, 1328) = 1.4$, $p = .02$.

Univariate tests were analyzed to further investigate the individual relationship each IV had on the DVs separately. Sphericity was violated for Instrument on roughness, $\chi^2(14) = 35.1$, $p = .002$ and on blend, $\chi^2(14) = 24.0$, $p = .05$. It was also violated for Interval on roughness, $\chi^2(27) = 94.2$, $p < .001$ and on blend, $\chi^2(27) = 63.0$, $p < .001$. Greenhouse-Geisser corrections of $\varepsilon = .63$ were used for Instrument on roughness, $\varepsilon = .66$ for Instrument on blend, $\varepsilon = .38$ for Interval on roughness and $\varepsilon = .48$ for Interval on blend. Sphericity values were unobtainable for Instrument*Interval, so Greenhouse-Geisser corrections of $\varepsilon = .29$ was used for Instrument*Interval on roughness and $\varepsilon = .27$ for Instrument*Interval on blend. There was a main effect of Instrument on both roughness, $F(3.18, 60.37) = 21.4$, $p < .001$, and blend, $F(3.29, 62.49) = 6.2$, $p < .001$. The main effect of Interval was also significant for both roughness, $F(2.65, 50.39) = 33.9$, $p < .001$ and blend, $F(3.37, 64.13) = 12.9$, $p < .001$. The interaction effect of Instrument*Interval was not significant for either roughness or blend under the Greenhouse-Geisser correction.

Bonferroni-corrected pairwise comparisons for the main effect of Instrument on roughness showed that HC/HC was rated significantly smoother than TR/TR and rougher than VB/VB. FL/FL was also rated significantly smoother than TR/TR and rougher than VB/VB and CL/CL. TR/TR was rated as significantly rougher than every other baseline instrument pair. This is paralleled by the fact that there was a unanimous consensus on the roughness of the trumpet in subject interviews conducted after each experiment session. VB/VB, on the other hand, was rated significantly less rough compared to every other baseline instrument pair. This poses no surprise, as the vibraphone stimuli used in the experiment had very few partials with a sound quality similar to that of pure tones.

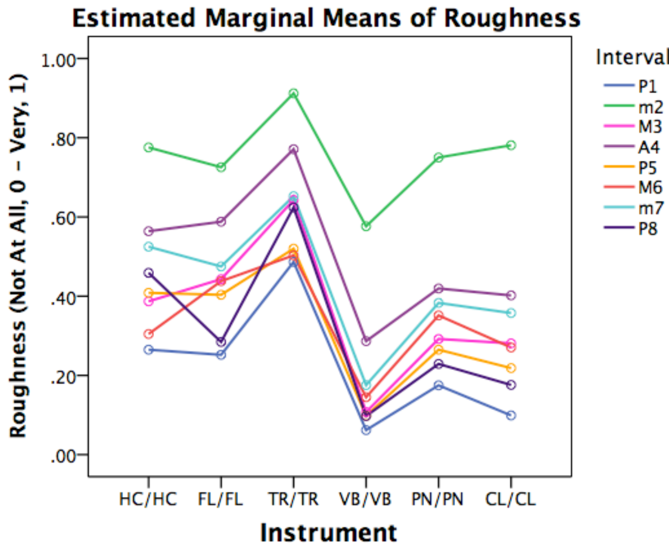

Fig. 4.2: Baseline roughness means for Interval as a function of Instrument.

The main effect of Instrument on blend showed that there were far fewer significant differences between the blend ratings of different instrument pairs, which were found only in relation to the flute. FL/FL was rated significantly less blended compared to VB/VB, PN/PN and CL/CL, but it was not significantly different from TR/TR or HC/HC.

The main effect of Interval on roughness was as expected. Interval P1 was rated significantly less rough than most other intervals, including m2, M3, A4 and m7. Interval m2 was rated significantly rougher than all other intervals with p-values of $< .001$ across the board. M3 was rated significantly rougher than P1, and significantly smoother than m2 and A4. A4 was rated significantly rougher than all other intervals except m2 and m7. P5 was rated significantly less rough compared to m2, A4 and m7. M6 was rated significantly less rough compared to m2, A4 and m7. Interval m7 was rated as rougher than all other intervals except m2 and A4. P8 was rated significantly less rough compared to m2, A4 and m7.

The main effect of Interval on blend was also generally as expected. All significant differences were in relation to intervals P1, m7 and P8. P1 was rated as more blended than all other intervals. Interval m7 was rated significantly less blended compared to P1, M3, P5, M6 and P8. Other than being rated significantly less blended compared to P1, P8 was rated significantly more blended compared to A4 and m7.

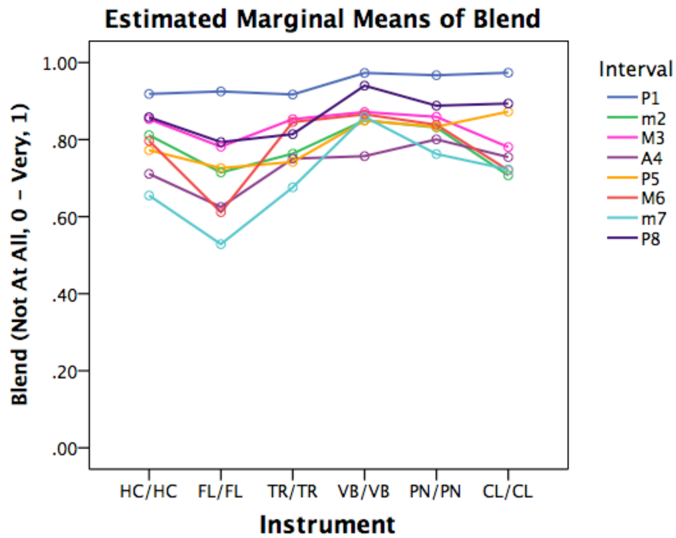

Fig. 4.3: Baseline blend means for Interval as a function of Instrument.

4.2.2 Analysis of Baseline Means for Pleasantness

Sphericity was violated for Instrument, $\chi^2(14) = 28.6$, $p = .01$ and Interval, $\chi^2(27) = 82.3$, $p < .001$. Greenhouse-Geisser corrections of $\varepsilon = .67$ and $\varepsilon = .35$ were used, respectively. Sphericity values were unobtainable for Instrument*Interval, so the Greenhouse-Geisser correction of $\varepsilon = .29$ was used. The main effects of Instrument, $F(3.36, 63.79) = 22.5$, $p < .001$ and Interval, $F(2.45, 46.53) = 8.3$, $p < .001$, were significant. The interaction effect of Instrument*Interval was not significant. The main effect of Instrument on pleasantness showed that all significant differences were in relation to baseline pairs TR/TR, VB/VB and CL/CL. TR/TR was rated as significantly less pleasant compared to HC/HC, VB/VB, PN/PN and CL/CL. This is corroborated by post-experiment comments made by the participants, whereby most stated that the trumpet stimuli were quite unpleasant. VB/VB was rated as significantly more pleasant than every other baseline pair. CL/CL was rated as significantly more pleasant than FL/FL and TR/TR, although it was rated significantly less pleasant compared to VB/VB.

The main effect of Interval on pleasantness showed that P1 and P8 were the only intervals that did not significantly differ from any other intervals on pleasantness. Significant differences for intervals m2, A4 and m7 showed lower pleasantness ratings, whereas significant differences for intervals M3, P5 and M6 showed higher pleasantness ratings. Interval m2 and A4 were rated significantly less pleasant compared to M3 and P5. Interval m7 was rated significantly less pleasant compared to M3, P5 and M6.

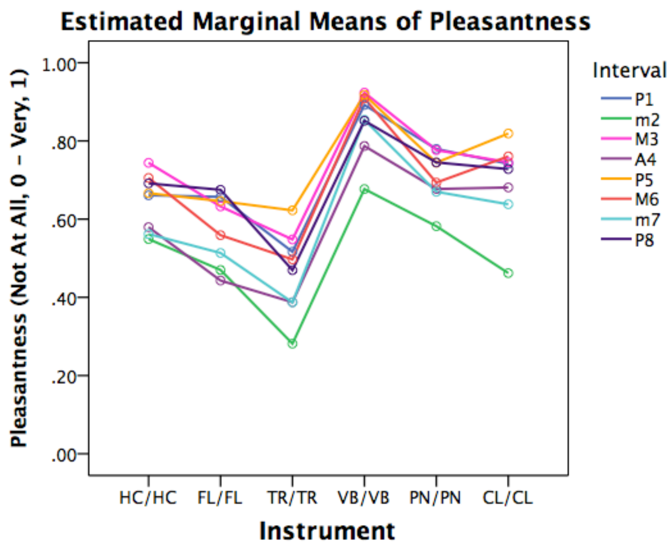

Fig. 4.4: Baseline pleasantness means for Interval as a function of Instrument.

4.2.3 Analysis of Experimental Means for Roughness and Blend

The combination of roughness and blend had significant multivariate effects on Pair, Wilk's $\lambda = .18$, $F(30, 568) = 25.5$, $p < .001$; Interval, Wilk's $\lambda = .2$, $F(14, 264) = 23.7$, $p < .001$; and Pair*Interval, Wilk's $\lambda = .83$, $F(210, 3988) = 1.9$, $p < .001$. Univariate tests were analyzed to further investigate the individual relationship each IV had on the DVs separately. Sphericity was violated for Pair on roughness, $\chi^2(119) = 206.1$, $p < .001$ and blend, $\chi^2(119) = 167.2$, $p = .01$, so the Greenhouse-Geisser corrections of $\varepsilon = .37$ and $\varepsilon = .42$ were used, respectively. Sphericity was violated for Interval on roughness, $\chi^2(27) = 94.6$, $p < .001$ and blend, $\chi^2(27) = 53.5$, $p = .002$. Greenhouse-Geisser corrections of $\varepsilon = .36$ and $\varepsilon = .53$ were used. Sphericity values were unobtainable for Pair*Interval, so the Greenhouse-Geisser corrections of $\varepsilon = .12$ was used for roughness and $\varepsilon = .14$ was used for blend. All main and interaction effects were significant for roughness and blend ratings. The main effect of Pair was significant for both roughness, $F(5.48, 104.18) = 15.7$, $p < .001$ and blend, $F(6.37, 121.12) = 41.5$, $p < .001$. The main effect of Interval was significant for roughness, $F(2.51, 47.68) = 26.6$, $p < .001$ and blend, $F(3.68, 70.0) = 32.1$, $p < .001$. The interaction effect of Pair*Interval was significant for roughness, $F(12.5, 237.55) = 1.9$, $p = .03$ and blend, $F(14.52, 275.79) = 1.9$, $p = .02$.

The main effect of Pair on roughness showed significant differences in relation to pairs TR/HC, HC/TR, TR/FL, FL/TR, VB/HC, HC/VB, VB/FL and HC/FL. In general, pairs including the trumpet tended to be rated significantly rougher than pairs without, whereas pairs including the vibraphone had the opposite effect. Baseline results for TR/TR and VB/VB support the accentuation and mitigation of roughness these two instruments had on different timbre combinations. The trumpet pairs (excluding the ones paired with the vibraphone) did not show any significant differences between themselves, suggesting that the trumpet's timbre contributed the most to a pair's overall roughness. Significant differences for pairs FL/CL and CL/FL were always in relation to their smoother quality compared to other pairs. Although the harpsichord was the strongest predictor of the tuning effects found in Experiments 1 and 2, harpsichord pairs were more affected by their instrument counterparts in terms of roughness perception in this experiment. There were no significant differences between any of the matched instrument pairs with switched order (e.g., TR/HC and HC/TR) in roughness, meaning that low note or high note positions did not notably affect the roughness of the overall timbre.

The main effect of Pair on blend showed that CL/FL, FL/CL, HC/PN, PN/HC, VB/HC, FL/TR and TR/FL tended to be rated higher in blend compared to other instrument pairs, whereas FL/HC, HC/FL, VB/TR and TR/VB tended to be rated as lower in blend. Overall, playing mechanism seemed to be the greatest predictor for blend. Pairs that were rated higher in blend were similar in playing mechanism (either both impulsive or both sustained), whereas pairs rated lower in blend had instruments with different playing mechanisms. There were two instances of a significant order difference between matched pairs for blend, which occurred between VB/HC and HC/VB, and FL/CL and CL/FL. VB/HC was rated as more blended than HC/VB, whereas CL/FL was rated as more blended than FL/CL.

The main effect of Interval on roughness and blend showed that results for roughness were unsurprisingly similar to that obtained for baseline measures. Averaged across all 16 pairs, m2 was rated significantly rougher than all other intervals, and A4 was significantly rougher than all other intervals excluding m2 and m7. Interval m7 was rated rougher than perfect intervals (P1, P5, P8), but was not significantly different from M3, A4 and M6. The perfect intervals were not significantly different from other musically consonant intervals, but were rated significantly smoother compared to musically dissonant intervals. M3 and M6 were not significantly different from musically consonant intervals as well as m7. The lack of difference between M3, M6 and m7 reflects on the ambiguous nature of the interval m7, since it is often not considered as strongly "dissonant" (in music theory terms) as other musically dissonant intervals like the minor second. Results for blend were somewhat typical. P1 was rated significantly more blended than all other pairs, whereas P8 was rated as significantly more blended than all other pairs except P1. Interval m7 was rated as significantly less blended compared to all musically consonant intervals except M6. What is interesting to note is that, averaged across all pairs, the blend rating for interval m2 was not significantly different from A4, P5, M6 and m7. M3 was rated significantly smoother than only the musically dissonant intervals (m2, A4, m7). A4 was not significantly different from m2, M6, or m7, whereas M6 was not significantly different from any other interval other than P1 and P8. Another interesting observation is that P5 was not significantly different from m2, M3 and M6.

An in-depth analysis of the interaction between Pair and Interval showed that, although the accentuation and mitigation of roughness happened at all intervals, it occurred in relation to different sets of pairs at certain intervals. Overall, there were more significant differences demonstrating roughness accentuation at musically consonant intervals, whereas there were more significant differences demonstrating roughness mitigation at musically dissonant intervals.

At P1, all of the significant differences were in relation to pairs FL/CL, TR/HC, HC/TR and HC/FL. In comparison to other pairs, FL/CL was rated significantly less rough, whereas TR/HC, HC/TR and HC/FL were rated significantly rougher. At m2, all significant differences were in relation to pairs VB/FL, VB/HC and TR/FL. In comparison to other pairs, VB/FL and VB/HC were rated significantly less rough, and TR/FL was rated significantly rougher. At M3, all significant differences were in relation to pairs VB/HC, FL/TR and HC/TR. In comparison to other pairs, VB/HC was rated significantly less rough, and FL/TR and HC/TR were rated significantly rougher. At A4, all significant differences were in relation to pairs TR/HC, FL/TR, HC/VB and VB/FL. In comparison to other pairs, HC/VB and VB/FL were rated significantly less rough, whereas TR/HC and FL/TR were rated significantly rougher. The only significant difference at P5 was between pairs TR/FL and VB/HC. TR/FL was rated significantly rougher compared to VB/HC. At M6, all significant differences were in relation to pairs TR/FL, FL/TR, VB/HC, VB/FL and HC/FL. In comparison to other pairs, VB/FL and VB/HC were rated significantly less rough, whereas FL/TR, TR/FL and HC/FL were rated significantly rougher. All significant differences at m7 were in relation to pairs TR/FL and VB/FL. In comparison to other pairs, VB/FL was rated significantly less rough, whereas TR/FL was rated significantly rougher. All significant differences at P8 were in relation to pairs TR/HC, HC/TR, HC/FL and FL/CL. In comparison to other pairs, FL/CL was rated significantly less rough, whereas TR/HC, HC/TR and HC/FL were rated significantly rougher.

As found in the main effect of Pair on roughness, the six trumpet pairs were always rated significantly rougher than pairs without the trumpet. The six vibraphone pairs were rated significantly less rough compared to pairs without the vibraphone, except when the vibraphone was paired with the trumpet and when compared to pairs FL/CL and CL/FL. Although FL/TR was rated as significantly rougher than TR/VB at M3, the six trumpet pairs otherwise did not significantly differ from each other. Pairs FL/CL and CL/FL were always rated less rough compared to other pairs at every interval except m2, where FL/CL was rated significantly rougher than VB/HC. There were no significant order differences between matched pairs for roughness at any interval.

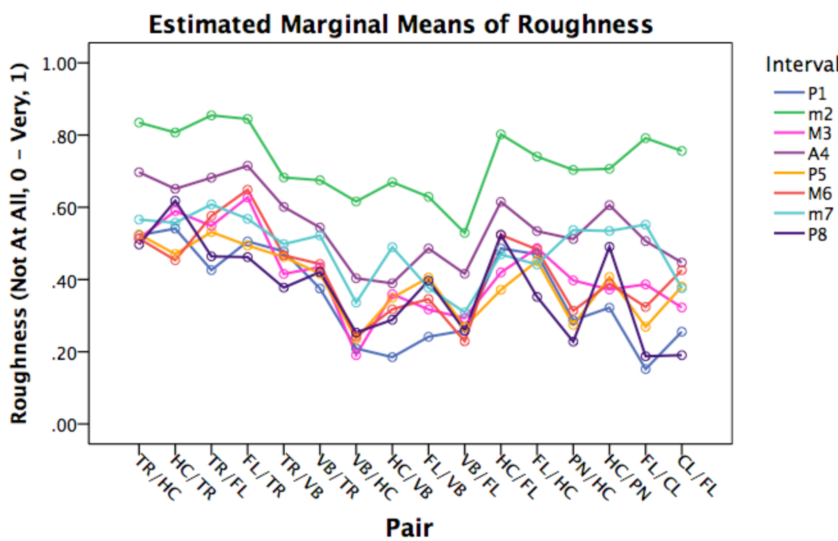

Fig. 4.5: Experimental roughness means for Interval as a function of Pair.

Significant roughness differences between intervals within each pair were unsurprising. For pairs TR/HC, TR/FL, VB/TR, VB/HC, FL/VB, VB/FL, FL/HC and CL/FL, significant differences between intervals occurred in relation to m2. For pairs HC/VB, HC/FL, PN/HC and HC/PN, significant differences between intervals occurred in relation to m2, A4 and m7. A4 and m7 were rated as significantly rougher than other intervals except m2, which was consistently rated as the roughest interval. Pair TR/VB did not show any significant differences between intervals. For HC/TR, m2 was rated significantly rougher than most other intervals, whereas M3 was rated significantly rougher than M6. For FL/TR, m2 was rated significantly rougher than most other intervals, A4 was rated significantly rougher than P5 and P8, and P8 was rated significantly smoother than m2, M3 and A4. For FL/CL, P1 was rated significantly smoother than most other intervals, m2 was rated significantly rougher than all other intervals, and m7 was rated rougher than P1, P5 and P8.

The interaction effect between Pair and Interval for blend showed significant increases and decreases in blend at all intervals, albeit in relation to different sets of pairs. Although there were significant differences demonstrating both increased and decreased blend at musically consonant intervals, there were more significant differences demonstrating increased blend at musically dissonant intervals. At P1, TR/HC, HC/TR, HC/FL and FL/HC were rated significantly less blended compared to other pairs, whereas VB/HC, HC/PN, FL/CL and CL/FL were rated significantly more blended. At m2, FL/HC was rated as significantly less blended compared to other pairs, whereas PN/HC, HC/PN, TR/FL and CL/FL were rated significantly more blended. At M3, TR/HC, HC/TR, TR/VB, VB/TR, HC/FL and FL/HC were rated significantly less blended compared to other pairs, whereas TR/FL and CL/FL were rated significantly more blended. At A4, TR/VB was rated significantly less blended compared to other pairs, whereas VB/HC, PN/HC, HC/PN and CL/FL were rated significantly more blended. At P5, FL/TR, VB/HC, PN/HC, HC/PN, FL/CL and CL/FL were rated significantly more blended compared to other pairs, whereas HC/FL was rated significantly less blended. At M6, PN/HC, HC/PN, FL/CL and CL/FL were rated significantly more blended compared to other pairs, whereas TR/VB and FL/HC were rated significantly less blended. At m7, PN/HC and CL/FL were rated as significantly more blended compared to other pairs. At P8, PN/HC, FL/CL and CL/FL were rated as significantly more blended compared to other pairs, whereas HC/FL and TR/HC were rated as significantly less blended.

There were no significant order differences between matched pairs at any interval except between pairs VB/HC and HC/VB at A4, where VB/HC was rated as significantly more blended than HC/VB. As found previously in the main effect of pair on blend, playing mechanism had the greatest effect on blend ratings. Regardless of the interval, pairs PN/HC, HC/PN, FL/CL and CL/FL with similar playing mechanisms were comparatively the most blended, whereas pairs

TR/HC, HC/TR, TR/VB, VB/TR, HC/FL and FL/HC with different playing mechanisms were comparatively the least blended. Pairs VB/HC, HC/VB, TR/FL and FL/TR were rated as more blended when compared to pairs that had instruments with different playing mechanisms.

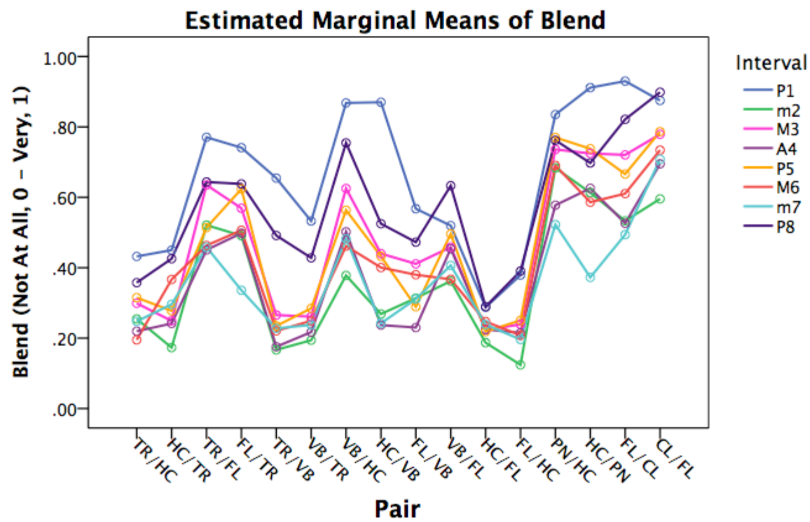

Fig. 4.6: Experimental blend means for Interval as a function of Pair.

Generally, musically dissonant intervals were rated as significantly less blended compared to musically consonant intervals, and vice versa. Pairs TR/VB, VB/HC, HC/VB, HC/PN and FL/CL had far more pronounced significant differences between intervals within themselves, whereas pairs VB/FL and HC/FL had no significant differences between intervals at all. For TR/VB, P1 was significantly more blended than all other intervals except P8, and P8 was significantly more blended than all other intervals except P1. For VB/HC, P1 was significantly more blended than all other intervals except M3 and P8, whereas m2 was also significantly less blended than P8. For HC/VB, P1 was significantly more blended than all other intervals, and A4 was significantly less blended than P8. For HC/PN, P1 was significantly more blended than all other intervals except P5, and m7 was significantly less blended than P1, M3, P5 and P8. For FL/CL, P1 was significantly more blended than all other intervals except P8, and m7 was significantly less blended than P1, M3 and P8.

4.2.4 Analysis of Experimental Means for Pleasantness

Sphericity was violated for Pair, $\chi^2(119) = 249.6$, $p < .001$ and Interval, $\chi^2(27) = 113.8$, $p < .001$. Greenhouse-Geisser corrections of $\varepsilon = .32$ and $\varepsilon = .34$ were used, respectively. Sphericity values were unobtainable for Pair*Interval, so the Greenhouse-Geisser correction of $\varepsilon = .12$ was used. The main effects of Pair, $F(4.87, 92.52) = 16.6$, $p < .001$ and Interval, $F(2.36, 44.84) = 10.0$, $p < .001$, were significant. The interaction effect of Pair*Interval was not significant under the Greenhouse-Geisser correction. For the main effect of Pair on pleasantness, significant differences in pleasantness were seen in relation to pairs VB/HC, HC/VB, VB/FL, FL/HC, PN/HC, HC/PN, CL/FL and HC/TR. In comparison to other pairs, VB/HC, HC/VB, VB/FL, PN/HC, HC/PN and CL/FL were rated significantly more pleasant, whereas FL/HC and HC/TR were rated significantly less pleasant. As seen for baseline means, pairs with the trumpet tended to receive lower pleasantness ratings and pairs with the vibraphone and clarinet tended to receive higher pleasantness ratings, suggesting a negative relation to roughness ratings. Similar to the trend seen for blend, pairs with similar playing mechanisms were somewhat more likely to be rated as more pleasant compared to pairs with different playing mechanisms. There were no order differences between matched pairs, meaning that high or low note position did not contribute to the overall pleasantness of an instrument pair.

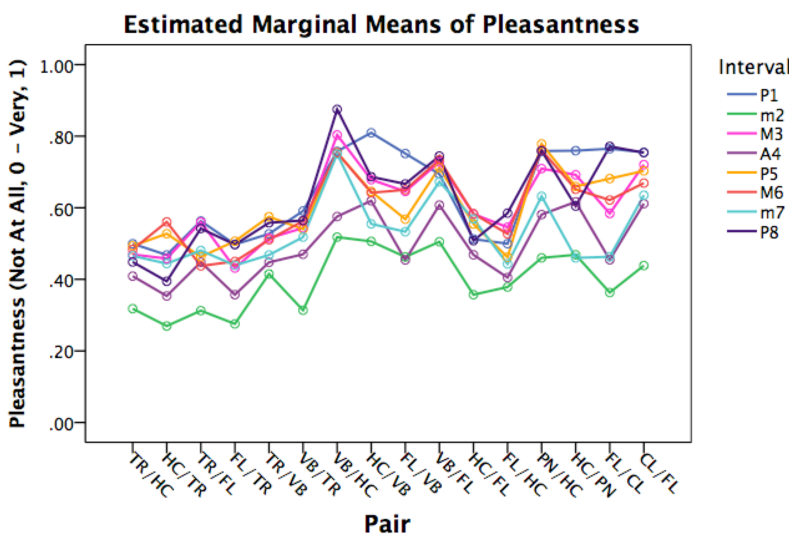

Fig. 4.7: Experimental pleasantness means for Interval as a function of Pair.

The main effect of Interval on pleasantness showed that significant differences only occurred in relation to m2 and A4. Interval m2 was significantly less pleasant than every other interval except M3, A4 and m7.

4.2.5 Comparison of Baseline and Experimental Means for Roughness and Blend

Doubly multivariate repeated-measures ANOVAs were run separately for each set of matched baseline and experimental pairs. The six baseline pairs were only compared to the experimental pairs that had the same instrument in either the low or high note position. That is, HC/HC was compared to TR/HC, HC/TR, VB/HC, HC/VB, HC/FL, FL/HC, PN/HC and HC/PN. FL/FL was compared to TR/FL, FL/TR, FL/VB, VB/FL, HC/FL, FL/HC, FL/CL and CL/FL. TR/TR was compared to TR/HC, HC/TR, TR/FL, FL/TR, TR/VB and VB/TR. VB/VB was compared to TR/VB, VB/TR, VB/HC, HC/VB, FL/VB and VB/FL. PN/PN was compared to PN/HC and HC/PN, and CL/CL was compared to FL/CL and CL/FL. All pairwise comparisons were corrected under a Bonferroni adjustment.

A 9 Pair x 8 Interval ANOVA was conducted to investigate significant differences in relation to HC/HC. The combination of roughness and blend had multivariate effects on Pair, Wilk's $\lambda = .13$, $F(16, 302) = 34.0$, $p < .001$; Interval, Wilk's $\lambda = .26$, $F(14, 264) = 17.8$, $p < .001$; and Pair*Interval, Wilk's $\lambda = .8$, $F(112, 2126) = 2.2$, $p < .001$. Univariate effects for Pair*Interval were also significant under the Greenhouse-Geisser correction of $\varepsilon = .2$ for roughness and $\varepsilon = .23$ for blend; $F(11.14, 211.76) = 1.9$, $p = .04$ and $F(13.01, 247.25) = 2.6$, $p = .002$, respectively. Bonferroni-corrected pairwise comparisons for roughness showed significant differences between HC/HC and other harpsichord pairs only at intervals P1, M3 and P8. At P1, HC/HC was rated less rough compared to TR/HC and HC/FL. At M3, HC/HC was rated less rough compared to HC/TR. At P8, HC/HC was rated rougher than PN/HC.

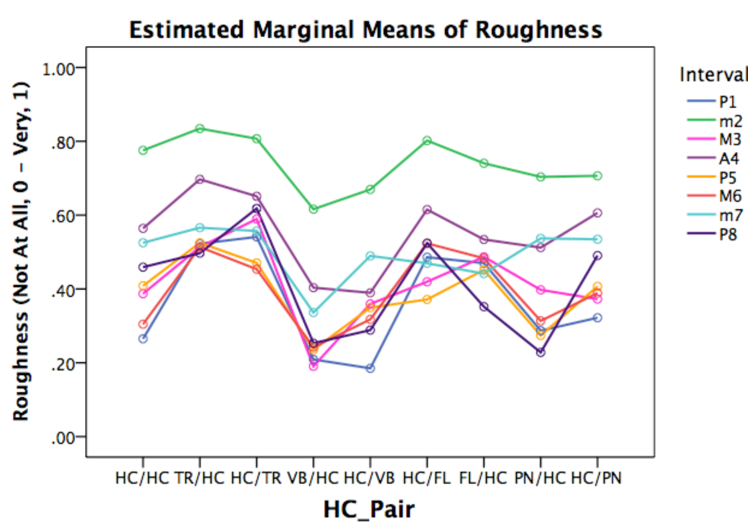

Fig. 4.8: Baseline and experimental roughness means for Interval as a function of HC_Pair.

Bonferroni-corrected pairwise corrections for blend showed significant differences between HC/HC and most other harpsichord pairs at every interval, which was to be expected. Although there is nothing surprising about the baseline pair of HC/HC being rated as more blended than experimental pairs, it is interesting to note that HC/HC blend ratings never significantly differed from PN/HC and usually did not differ from VB/HC and HC/PN. At P1 and P8, HC/HC was rated significantly more blended compared to TR/HC, HC/TR, HC/FL and FL/HC, but not VB/HC, HC/VB, PN/HC, or HC/PN. At m2, HC/HC was rated significantly more blended compared to all other harpsichord pairs except PN/HC and HC/PN. At M3, A4 and P5, HC/HC was significantly more blended compared to all other harpsichord pairs except VB/HC, PN/HC and HC/PN. At M6, HC/HC was significantly more blended compared to all other harpsichord pairs except PN/HC and HC/PN. At m7, HC/HC was significantly more blended compared to all other harpsichord pairs except VB/HC and PN/HC.

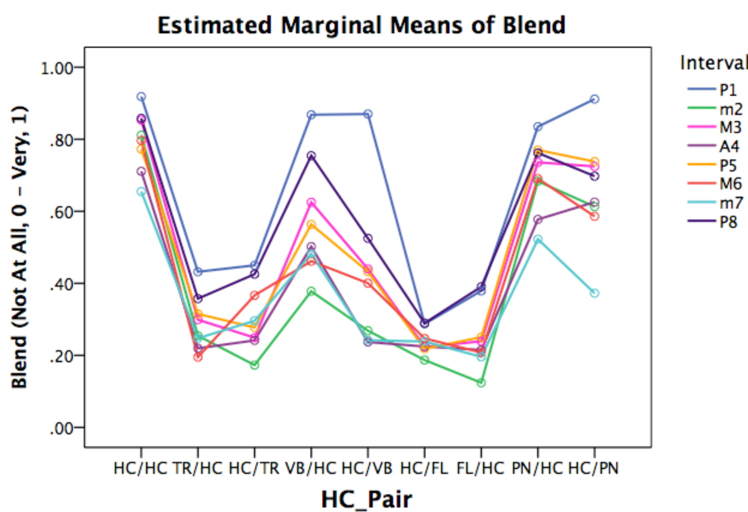

Fig. 4.9: Baseline and experimental blend means for Interval as a function of HC_Pair.

A 9 Pair x 8 Interval ANOVA was conducted to investigate significant differences in relation to FL/FL. The combination of roughness and blend had multivariate effects on Pair, Wilk's $\lambda = .16$, $F(16, 302) = 28.4$, $p < .001$, Interval, Wilk's $\lambda = .26$, $F(14, 264) = 18.1$, $p < .001$ and Pair*Interval, Wilk's $\lambda = .84$, $F(112, 2126) = 1.7$, $p < .001$. The univariate effect for Pair*Interval was also significant under the Greenhouse-Geisser correction of $\varepsilon = .2$ for roughness, $F(11.16, 212.12) = 2.0$, $p = .03$. However, Bonferroni-corrected pairwise comparisons for roughness showed that there were no significant differences between FL/FL and other flute pairs at any interval.

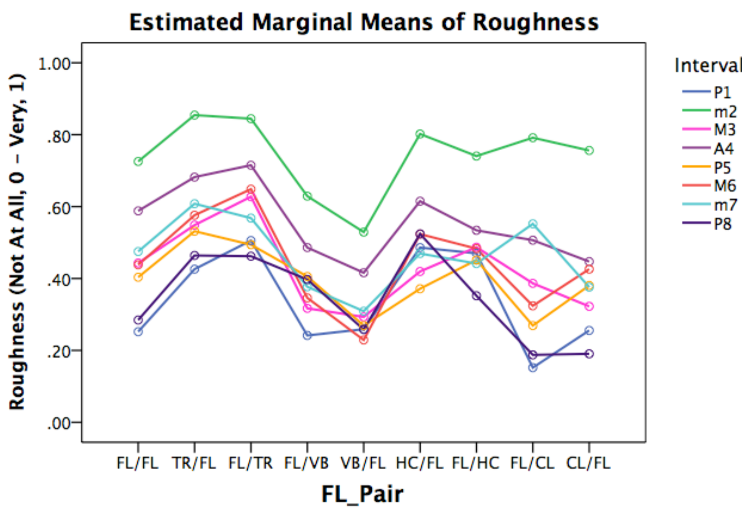

Fig. 4.10: Baseline and experimental roughness means for Interval as a function of FL_Pair.

Despite the fact that the univariate Pair*Interval interaction effect was not significant for blend ratings of flute pairs after the Greenhouse-Geisser correction of $\varepsilon = .23$, there were significant differences in blend ratings for the main effect of Pair in relation to FL/FL, $F(3.76, 71.47) = 42.2$, $p < .001$ (corrected using $\varepsilon = .47$). Averaged across all intervals, FL/FL was rated as significantly more blended compared to all other flute pairs except FL/CL and CL/FL.

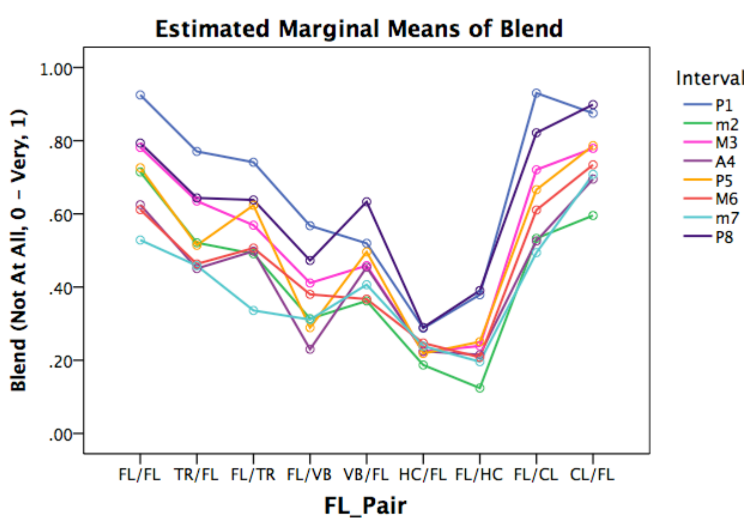

Fig. 4.11: Baseline and experimental blend means for Interval as a function of FL_Pair.

A 6 Pair x 8 Interval ANOVA was conducted to investigate significant differences in relation to TR/TR. The combination of roughness and blend had multivariate effects on Pair, Wilk's $\lambda = .19$, $F(12, 226) = 24.1$, $p < .001$, Interval, Wilk's $\lambda = .25$, $F(14, 264) = 18.5$, $p < .001$ and Pair*Interval, Wilk's $\lambda = .86$, $F(84, 1594) = 1.5$, $p = .003$. However, univariate effects for Pair*Interval were significant for neither roughness nor blend under the Greenhouse-Geisser corrections of $\varepsilon = .25$ and $\varepsilon = .27$, respectively. The univariate main effect of Pair for roughness, $F(4.0, 76.03) = 8.5$, $p < .001$ (corrected using $\varepsilon = .67$), showed that TR/TR was only significantly rougher compared to pairs TR/VB and VB/TR when averaged across all intervals.

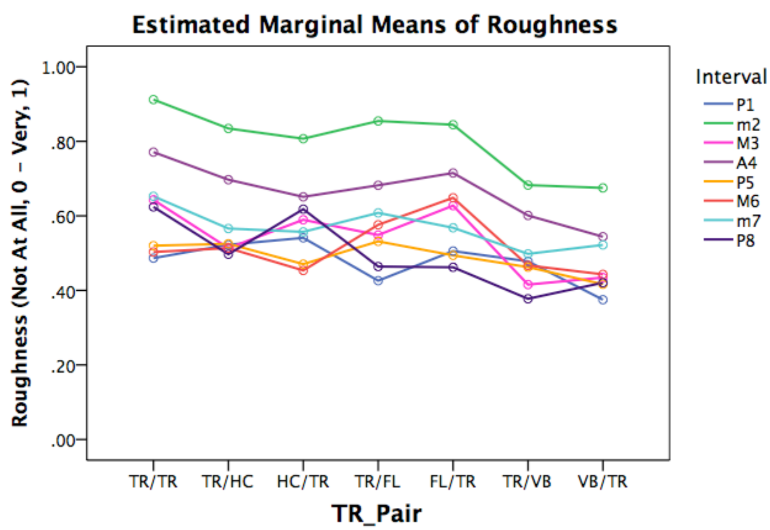

Fig. 4.12: Baseline and experimental roughness means for Interval as a function of TR_Pair.

The univariate main effect of Pair for blend, $F(3.06, 58.12) = 53.5$, $p < .001$ (corrected using $\varepsilon = .51$), showed that TR/TR was rated as significantly more blended than every other trumpet pair when averaged across all intervals, with p-values all < .001.

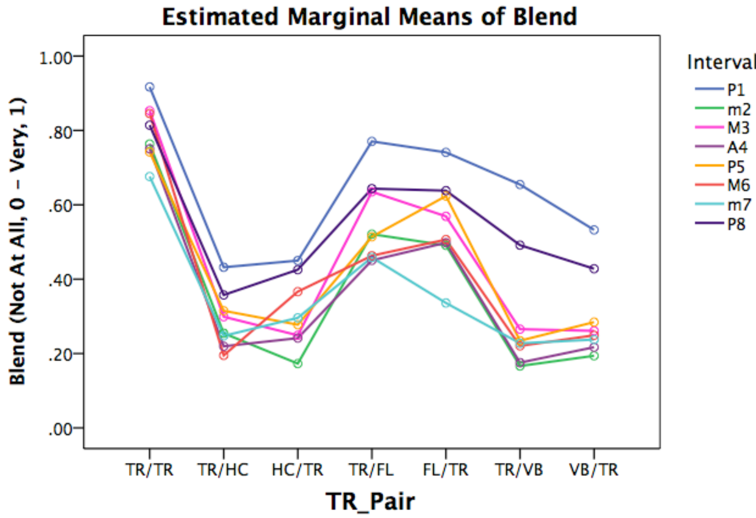

Fig. 4.13: Baseline and experimental blend means for Interval as a function of TR_Pair.

A 6 Pair x 8 Interval ANOVA was conducted to investigate significant differences in relation to VB/VB. The combination of roughness and blend had multivariate effects on Pair, Wilk's $\lambda = .21$, $F(12, 226) = 22.6$, $p < .001$, Interval, Wilk's $\lambda = .23$, $F(14, 264) = 20.3$, $p < .001$ and Pair*Interval, Wilk's $\lambda = .83$, $F(84, 1594) = 1.9$, $p < .001$. Although the univariate interaction effect of Pair*Interval for roughness was not significant after the Greenhouse-Geisser correction ($\varepsilon = .25$), the univariate main effect of Pair for roughness was significant after correction ($\varepsilon = .51$), $F(3.1, 58.7) = 17.4$, $p < .001$. Averaged across all intervals, VB/VB was rated as significantly less rough compared to all other vibraphone pairs.

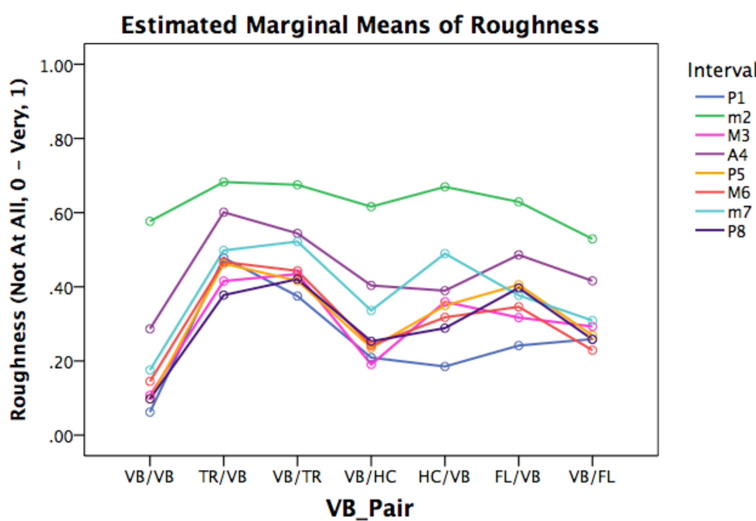

Fig. 4.14: Baseline and experimental roughness means for Interval as a function of VB_Pair.

The univariate interaction effect of Pair*Interval for blend was significant after correction (ε = .28), $F(11.76, 223.43) = 2.21$, $p = .01$. At P1, VB/VB was significantly more blended compared to all other vibraphone pairs except VB/HC and HC/VB. At m2, M6 and m7, VB/VB was significantly more blended compared to all other vibraphone pairs. At M3, A4, P5 and P8, VB/VB was significantly more blended compared to all other vibraphone pairs except VB/HC.

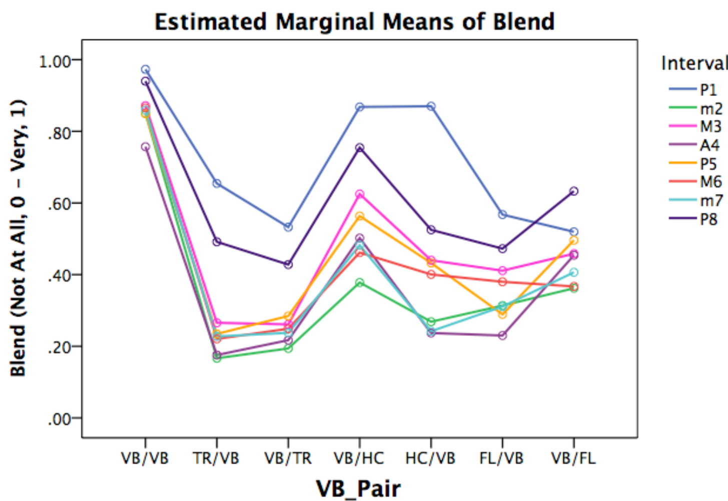

Fig. 4.15: Baseline and experimental blend means for Interval as a function of VB_Pair.

A 3 Pair x 8 Interval ANOVA was conducted to investigate significant differences in relation to PN/PN. The combination of roughness and blend had multivariate effects on Pair, Wilk's $\lambda = .41$, $F(4, 74) = 10.2$, $p < .001$, Interval, Wilk's $\lambda = .34$, $F(14, 264) = 13.6$, $p < .001$ and Pair*Interval, Wilk's $\lambda = .83$, $F(28, 530) = 1.9$, $p = .005$. Although the univariate interaction effect of Pair*Interval for roughness was not significant after the Greenhouse-Geisser correction (ε = .43), the univariate main effect of Pair for roughness was significant after correction (ε = .74), $F(1.48, 28.15) = 8.4$, $p = .003$. PN/PN was rated significantly less rough compared to HC/PN when averaged across all intervals.

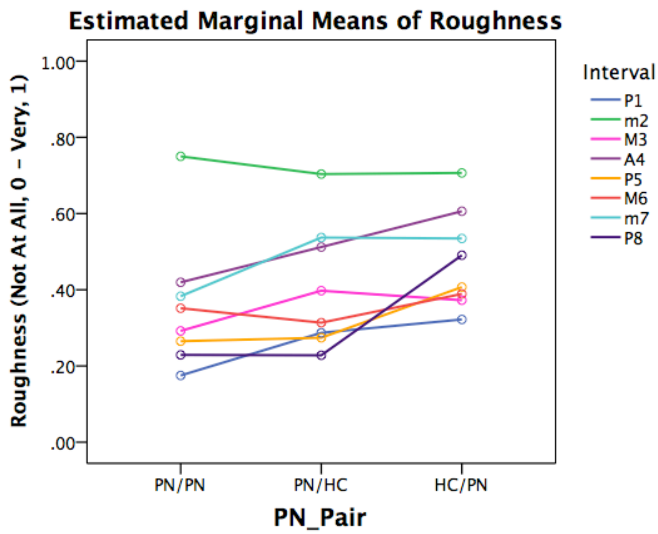

Fig. 4.16: Baseline and experimental roughness means for Interval as a function of PN_Pair.

The univariate interaction effect of Pair*Interval for blend was not significant, but the univariate main effect of Pair for blend was significant, $F(2, 38) = 21.5$, $p < .001$. PN/PN was rated significantly more blended than PN/HC and HC/PN when averaged across all intervals.

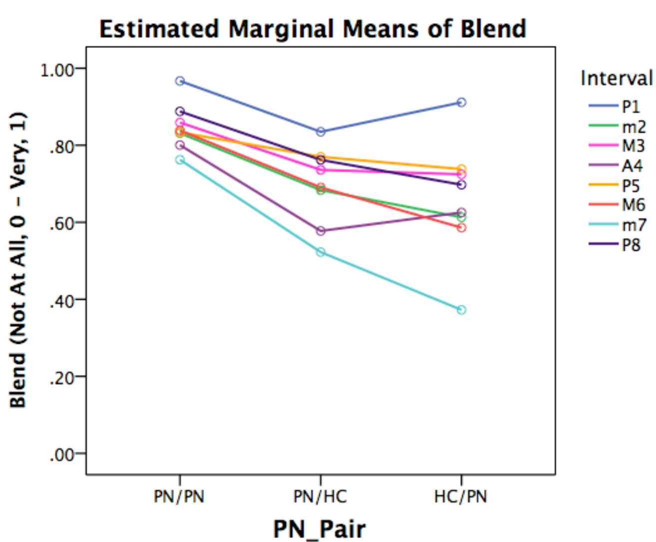

Fig. 4.17: Baseline and experimental blend means for Interval as a function of PN_Pair.

A 3 Pair x 8 Interval ANOVA was conducted to investigate significant differences in relation to CL/CL. The combination of roughness and blend had multivariate effects on Pair, Wilk's $\lambda = .3$, $F(4, 74) = 15.4$, $p < .001$, Interval, Wilk's $\lambda = .33$, $F(14, 264) = 13.9$, $p < .001$ and

Pair*Interval, Wilk's $\lambda = .84$, $F(28, 530) = 1.7$, $p = .02$. The univariate interaction effect of Pair*Interval was significant for roughness under the Greenhouse-Geisser correction of $\varepsilon = .54$, $F(7.49, 142.3) = 2.2$, $p = .03$, but not blend. At P1, P5 and M6, CL/CL was significantly less rough compared to CL/FL, whereas at m7, CL/CL was significantly less rough compared to FL/CL.

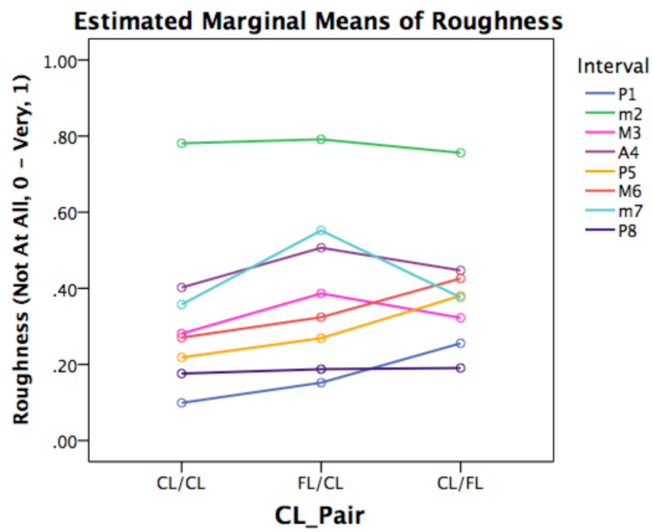

Fig. 4.18: Baseline and experimental roughness means for Interval as a function of CL_Pair.

The univariate main effect of Pair was significant for blend, $F(2, 38) = 21.8$, $p < .001$. Averaged across all intervals, CL/CL was only significantly more blended compared to FL/CL.

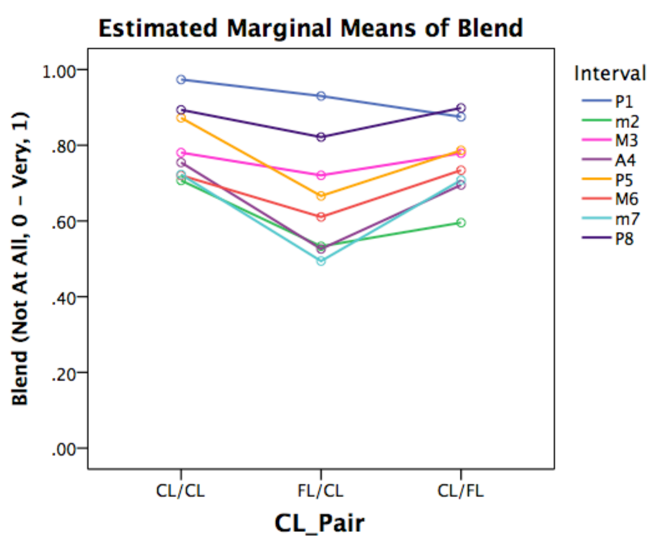

Fig. 4.19: Baseline and experimental blend means for Interval as a function of CL_Pair.

4.2.6 Comparison of Baseline and Experimental Means for Pleasantness

Univariate analyses for pleasantness ratings were run similarly to those for roughness and blend. Separate ANOVAs were conducted for each baseline pair along with their experimental counterparts. A 9 Pair x 8 Interval ANOVA for harpsichord pairs showed that the interaction effect of Pair*Interval was significant under the Greenhouse-Geisser correction of $\varepsilon = .2$, $F(12.02, 228.34) = 1.9$, $p = .03$. At m2 and A4, HC/HC was rated as significantly more pleasant compared to HC/TR. At M3, HC/HC was rated as significantly more pleasant than TR/HC and HC/TR. At m7, HC/HC was significantly less pleasant than VB/HC. At P8, HC/HC was rated more pleasant compared to HC/TR and HC/FL.

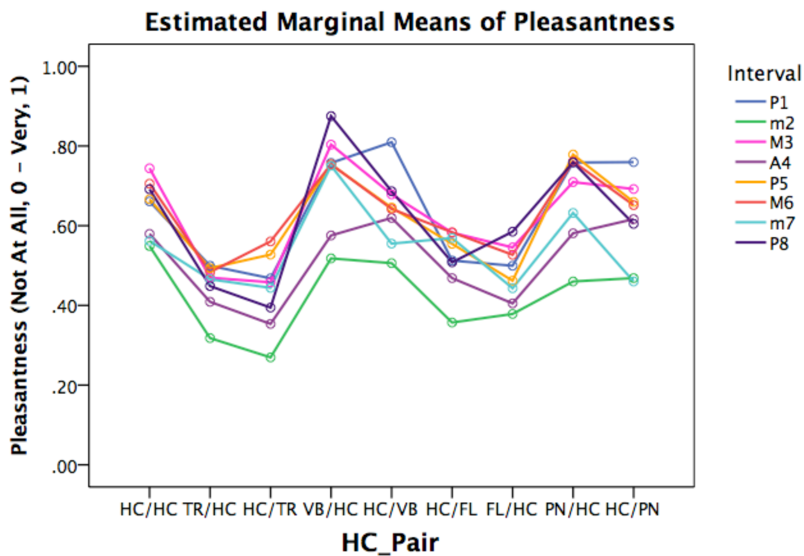

Fig. 4.20: Baseline and experimental pleasantness means for Interval as a function of HC_Pair.

A 9 Pair x 8 Interval ANOVA for flute pairs showed that the interaction effect of Pair*Interval was not significant. However, the main effect of Pair was significant under the Greenhouse-Geisser correction of $\varepsilon = .5$, $F(4.09, 77.76) = 15.2$, $p < .001$. Averaged across all intervals, FL/FL was rated more pleasant compared to FL/TR, but did not significantly differ from any other flute pair.

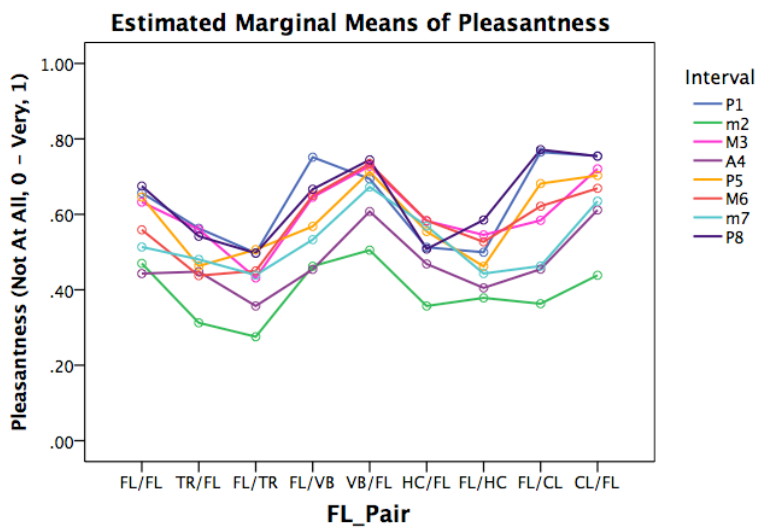

Fig. 4.21: Baseline and experimental pleasantness means for Interval as a function of FL_Pair.

A 6 Pair x 8 Interval ANOVA for trumpet pairs showed that neither the interaction effect of Pair*Interval nor the main effect of Pair were significant. For pairs that included the trumpet, instrumental differences did not significantly affect pleasantness ratings and only the effect of interval made a difference. This lack of difference between baseline and experimental means suggests that the timbre of the trumpet was the greatest contributor to the pleasantness of these pairs, regardless of the instrument playing the other note.

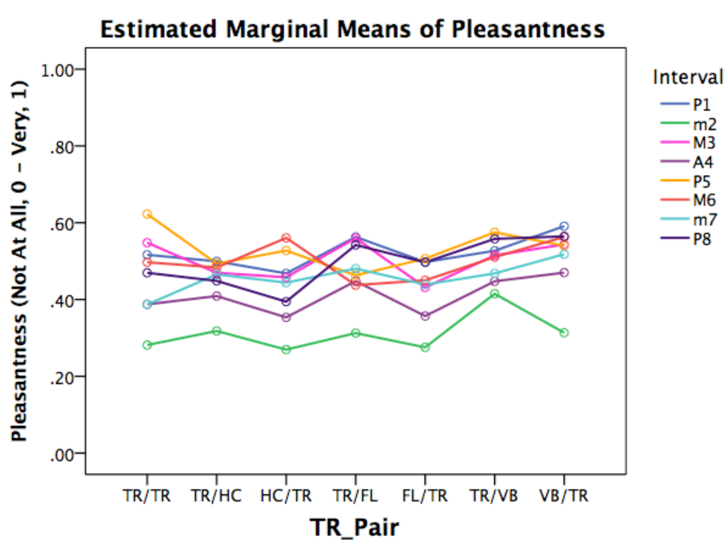

Fig. 4.22: Baseline and experimental pleasantness means for Interval as a function of TR_Pair.

A 6 Pair x 8 Interval ANOVA for vibraphone pairs showed that the interaction effect of Pair*Interval was not significant. However, the main effect of Pair was significant under the Greenhouse-Geisser correction of $\varepsilon = .5$, $F(2.99, 56.88) = 26.8$, $p < .001$. Averaged across all intervals, VB/VB was rated more pleasant than all other vibraphone pairs.

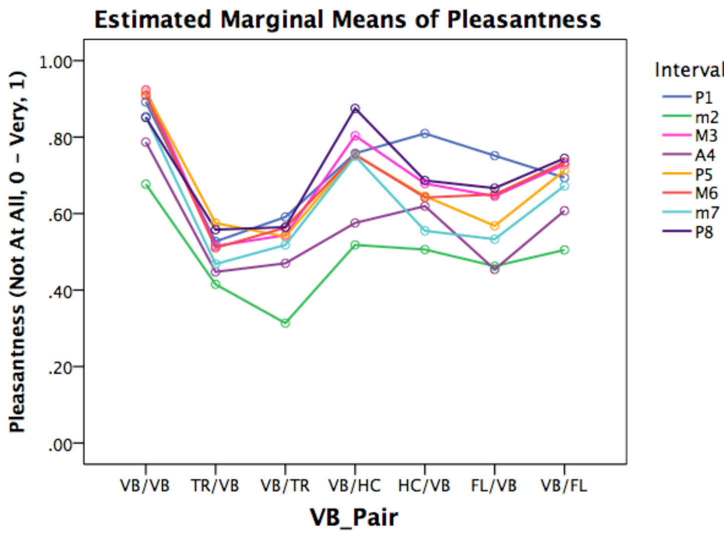

Fig. 4.23: Baseline and experimental pleasantness means for Interval as a function of VB_Pair.

A 3 Pair x 8 Interval ANOVA for piano pairs showed that the interaction effect of Pair*Interval was not significant. The main effect of Pair was significant, $F(2, 38) = 5.2$, $p = .01$. Averaged across all intervals, PN/PN was rated more pleasant compared to HC/PN.

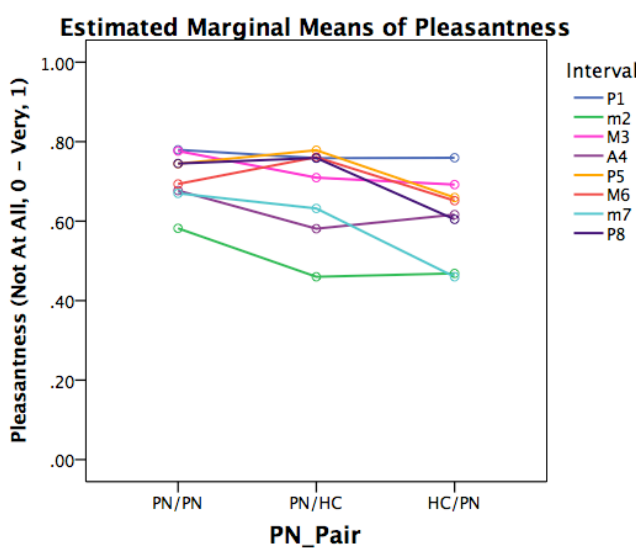

Fig. 4.24: Baseline and experimental pleasantness means for Interval as a function of PN_Pair.

A 3 Pair x 8 Interval ANOVA for clarinet pairs showed that the interaction effect of Pair*Interval was not significant. The main effect of Pair was significant, $F(2, 38) = 11.5$, $p < .001$. Averaged across all intervals, CL/CL was rated more pleasant compared to FL/CL.

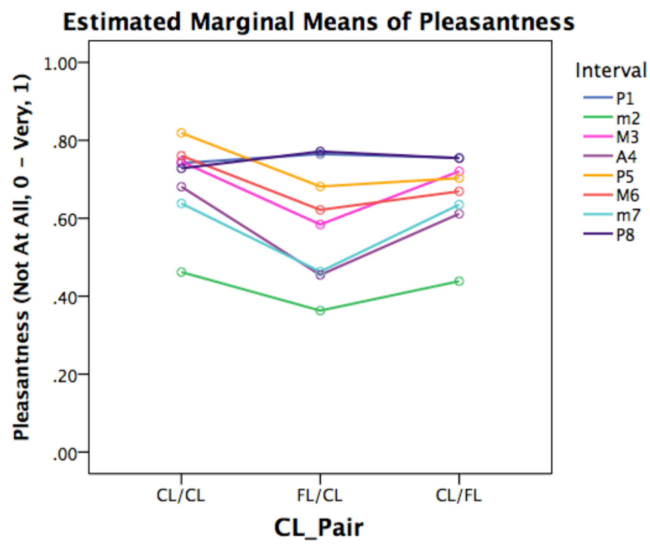

Fig. 4.25: Baseline and experimental pleasantness means for Interval as a function of CL_Pair.

4.3 Discussion

Results showed that for both baseline and experimental means, the trumpet had the strongest influence on accentuating roughness. Trumpet pairs were rated as similarly rough (not significantly different from each other) regardless of the instrument playing the other note, except for when the vibraphone played the other note. Trumpet pairings were also rated rougher than other pairings at all intervals and there were no order differences seen between matched trumpet pairs, suggesting that the trumpet carried the most weight in the overall timbre of pairs TR/HC, HC/TR, TR/FL, FL/TR, TR/VB and VB/TR. The vibraphone also strongly influenced the mitigation of roughness in vibraphone pairs at all intervals. In addition to the effects seen for the trumpet and vibraphone, pairs HC/FL and FL/HC were consistently rated rougher than other pairs (excluding trumpet pairs), whereas HC/PN, PN/HC, FL/CL and CL/FL were consistently rated smoother than other pairs (excluding vibraphone pairs). There were significant differences seen at all intervals, usually in relation to trumpet and vibraphone pairs, but roughness accentuation (more differences in relation to greater roughness) was more apparent at musically consonant intervals, whereas roughness mitigation (more differences in relation to greater smoothness) was more apparent at dissonant intervals. There were no order differences seen in any matched pair, suggesting that high or low note position did not affect the roughness of the overall timbre.

Comparisons between baseline and experimental means for roughness showed some interesting results. HC/HC was only significantly less rough than some of the other harpsichord pairs at P1 and M3, and was actually rated rougher than PN/HC at P8. TR/TR was rated rougher than TR/VB and VB/TR when averaged across all intervals. PN/PN did not significantly differ in roughness from PN/HC overall, and CL/CL was only significantly less rough compared to other clarinet pairs at P1, P5, M6 and m7. Only VB/VB was significantly smoother compared to other vibraphone pairs when averaged across all intervals, and FL/FL did not differ from any other flute pair at any interval. The roughness accentuation effect of the trumpet timbre (a bright, sustained instrument) is shown in the lack of difference between baseline and experimental trumpet pairs excluding the two with the vibraphone, whereas the roughness mitigation effect of the vibraphone (a dull, impulsive instrument) is shown in the significant differences between baseline and experimental vibraphone pairs. It is notable that certain instruments contributed much more to the perception of roughness than others, but it is unclear how much weight their timbral properties such as brightness or attack have in comparison to the mere perception of beating in the overall sound.

Blend ratings showed that participants based their judgements largely on playing mechanism for experimental means. Among pairs with similar playing mechanisms, CL/FL, FL/CL, PN/HC and HC/PN had the highest blend ratings, and among pairs with different playing mechanisms, FL/HC, HC/FL, TR/VB and VB/TR received the lowest blend ratings. Only matched pairs VB/HC and HC/VB showed an instance of order importance at A4, but there were otherwise no significant order differences between matched pairs at any interval. It is possible that different playing mechanisms prompted instability or incoherence in the composite timbres of instrument pairs, reducing blend (Brues, 1927; Culling & Darwin, 1993). However, for baseline means, the only significant differences between instrument pairs were in relation to FL/FL, which was significantly less blended compared to VB/VB, PN/PN and CL/CL; TR/TR and HC/HC were not significantly different from any other baseline pairs. Since all baseline pairs consisted of the same instrument playing both notes, playing mechanism was not the deciding factor. There is no clear explanation as to why participants found the flute less blended than the vibraphone, piano and clarinet, other than a tentative guess that the result was either a by-product of the sample size or the stimuli used in the experiment. A deeper analysis into the specific spectral envelopes of each stimuli combination at each interval may provide a more comprehensive picture of blend, as Lembke & McAdams (2015) have found that local, rather than global, spectral envelope descriptors may be more meaningful in explaining the blend of wind instrument dyads.

Although it was expected that baseline pairs would always receive higher blend ratings compared to experimental pairs regardless of the interval, comparisons between baseline and experimental means for blend showed otherwise. HC/HC was never rated more blended than PN/HC at any interval, and usually did not differ from VB/HC and HC/PN. FL/FL did not differ from FL/CL and CL/FL overall, VB/VB did not differ from VB/HC except at m2, M6 and m7, and CL/CL did not differ from CL/FL. Only TR/TR and PN/PN were significantly more blended than other trumpet and piano pairs.

Ratings for pleasantness strongly depended on the instrumental timbre of the trumpet and vibraphone, which was similar to what was found for roughness. Participants rated TR/TR as significantly less pleasant compared to all other baseline pairs except FL/FL, whereas VB/VB was rated more pleasant than all other baseline pairs. This finding was generally carried over to the experimental means, where trumpet pairs tended to receive low pleasantness ratings and vibraphone and clarinet pairs tended to receive the opposite. Pairs with similar playing mechanism were also more likely to be rated as pleasant compared to pairs with different playing mechanisms, but the timbre influences of the trumpet and vibraphone had greater effect on rating outcomes. There were also no order differences between matched pairs, suggesting that high or low note position did not affect the pleasantness of the overall timbre.

Comparisons between baseline and experimental means for pleasantness were also somewhat similar to what was found for roughness ratings. TR/TR was not significantly different from any other trumpet pair at any interval, whereas VB/VB was significantly more pleasant than all other vibraphone pairs overall. FL/FL was only more pleasant compared to FL/TR when averaged across all intervals, whereas PN/PN was only more pleasant compared to HC/PN and CL/CL was only more pleasant compared to FL/CL. HC/HC was more pleasant compared to TR/HC, HC/TR and HC/FL at m2, M3, A4 and P8, and less pleasant compared to VB/HC at m7.

Overall results for pleasantness suggest a negative linear correlation with roughness, meaning that participants tended to find rougher timbres less pleasant and smoother timbres more pleasant. As auditory roughness is a strong cue for sensory dissonance, this unsurprising finding has been recorded before (Costa et al., 2000).

Separate bivariate correlations comparing Experiment 3 means (averaged across all subjects) to the baseline and experimental means of Experiments 1 and 2 showed no significant linear relationships between tuning deviations and perceptual ratings of roughness, blend and pleasantness. Although participants consistently perceived the timbre of certain instrumental pairs in certain ways at different intervals, their perceptual ratings did not simply explain the tuning patterns that were only apparent at musically consonant intervals. However, Experiment 3 provided support for the blend assumptions made in Experiments 1 and 2. Indeed, participants rated pairs HC/PN, PN/HC, FL/TR and TR/FL among the most blended of all experimental pairs and rated pairs FL/HC, HC/FL, TR/VB and VB/TR among the least blended, just as their distances in Lakatos' (2000) combined timbre space suggested. Additionally, participants rated musically dissonant intervals significantly rougher in comparison to musically consonant intervals, supporting the idea that sensory dissonance is associated with greater tuning variability at dissonant intervals, and possibly just associated with dissonant intervals overall.

Chapter 5

Conclusion

In general, the musicians who participated in these studies did not stray greatly from equal temperament when tuning harmonic intervals regardless of instrument pair or tuning position, even when they were instructed to use their subjective judgement on evaluating whether or not an interval was "in-tune." This was likely due to the fact that all of the participants were enculturated to Western tonal scales, and past findings on intonation and tuning have shown that musicians deviate less from equal temperament compared to other tuning systems (Karrick, 1998; Kopiez, 2003). Moreover, when measured against Karrick's (1998) Table 1 showing directional cent deviation values from equal temperament for the Pythagorean and Just tuning systems (p. 117), the mean tuning deviation values of Experiments 1 and 2 were often closer to equal temperament (±0) at each interval. These averaged values do not closely adhere to any particular tuning system, largely due to the explicit instructions given prior to each experimental session to keep tunings subjective, but it is possible that these mean values are not capturing the systemic differences between certain types of subjects.

In Experiments 1 and 2, musicians were especially sensitive to instrument order when tuning harpsichord pairs. Harpsichord pairs showed strong instrument order differences when tuned at musically consonant intervals, prompting interval contraction if the harpsichord was playing the upper note and interval expansion if it was playing the lower note, irrespective of the instrument playing the note the participant was tuning. Experiments 1 and 2 showed that tuning accuracy in relation to equal temperament is not necessarily correlated with tuning difficulty. Vibraphone pairs, comparatively the least rough and most blended, were tuned the most accurately but also had the widest confidence intervals, suggesting greater tuning difficulty (which corresponds to participant statements). A likely explanation for this, however, is that the much simpler spectra of the vibraphone stimuli stood out as being rather different from the other, more complex sounds.

Although there were explicit order differences seen in the tuning experiments for certain matched pairs, the third experiment did not produce the same order differences in terms of perception. Matched pairs were perceived relatively similarly for roughness, blend and pleasantness, but tuning results differed significantly. Roughness and pleasantness ratings largely depended on the inclusion of the trumpet (rough, unpleasant) or vibraphone (smooth, pleasant), whereas blend ratings largely depended on whether or not instrument pairs had similar or varying playing mechanisms. Perception of the overall timbre quality of a harmonic interval does not seem to greatly affect tuning accuracy or variability, which may simply be due to the fact that perceptual judgements of a sound rely on its overall timbre whereas tuning a sound requires greater focus on the instrument being tuned. Perhaps this is why disjunctions between tuning and perception exist in musical performances, supporting Geringer and Witt's (1985) finding that there is a "limited degree of association between perceptual judgments and performance" (p. 97) in string players. Altogether, participants showed greater tuning variability—but not tuning inaccuracy—at musically dissonant intervals, which were also consistently rated as perceptually rougher than musically consonant intervals. Significant differences in tuning were more apparent at consonant intervals due to the lower tuning variability in comparison to dissonant intervals. Bidelman and Krishnan (2009) found in a brainstem frequency-following response study that non-musicians show preferential encoding for musically consonant intervals and more robust responses to them, suggesting greater pitch salience at consonant intervals. Although the experiments conducted for this thesis did not study non-musicians, participant responses were unanimous on the comparative ease of consonant interval tuning. It is possible that musicians find it easier to tune consonant intervals and to recognize whether or not a consonant interval is out-of-tune, especially within the context of Western music, in which musically dissonant intervals are generally not as common.

Some limitations exist due to the practical scope of this master's thesis. The participant pool was on the smaller side for all three experiments at N = 20; the repeated-measures design produced far more cases compared to variables, increasing the noise in the data. Although the data were analyzed mostly using repeated-measures ANOVA, multilevel model analyses may have been more suitable for the experimental designs used in these experiments. However, the dataset was unfortunately not appropriate for such complex nested analyses. Due to the duration of Experiment 2, there was little variation in bright/impulsive, bright/sustained, dull/impulsive and dull/sustained instruments to show results beyond those relating to the harpsichord. Given that the harpsichord was the only bright/impulsive instrument used in the study, it is hard to say if the effects found are particular to the instrument timbre itself or particular to the combination of its properties. Although the stimuli were edited as much as possible through pre-experiment perceptual testing to achieve similar onset synchrony between pairs, onsets were adjusted for each instrument using a global value for all of its pitches. This did not effectively remove the slight lag in the flute onsets when paired with the percussive instruments at certain intervals, potentially creating biases in terms of tuning and blend (i.e., lower onset synchrony decreases blend). The timeframe and scope of the experiments did not allow for more controlled designs to reduce residual biases from musicians' instrument-specific experiences, which were not accounted for when analyzing the data. Additionally, these results are particular to the octave range of C4 to C5 and the instruments chosen for the experiment, and cannot readily be generalized to other ranges and instruments due to the variability in timbres between different instruments and registers. The results are also particular to dyads, since sensory dissonance has been proven to fluctuate relative to musical context (Mashinter, 2006; Di Stefano & Bertolaso, 2014) and different tuning and perceptual results would be obtained in chords containing more than two notes.

Despite its limitations, the findings of this thesis are applicable to composers, conductors and instrumentalists in orchestral composition and performance. Although the overall timbre of a harmonic interval is most prevalent when evaluating their perceptual quality, musicians focus more on the specific timbre of the instrument being adjusted when tuning. This supports Di Stefano and Bertolaso's (2014) claim that sensory consonance needs to be looked at in a multidimensional way. Experiment 3 did not find any explicit correlations between the perceptual ratings and the tuning deviations of different timbre combinations at various harmonic intervals, but this only means that they are not linearly related. Some other non-linear relationship could exist between these perceptual ratings and tuning deviations that cannot be captured within the scope of this thesis due to the fact that there is a lack of musical context in the stimuli, and the addition of a "tonal melodic environment" may increase or change tuning precision (Rakowski, 1990, p. 70).

Enhanced knowledge about the effects of sensory dissonance on tuning and intonation can contribute to a more defined way of controlling it in orchestral performance, especially for instrumentalists. In line with Sethares' work on adaptive tuning scales (1994) and local consonance (1993), this research lends itself to the execution of advanced instrumental techniques through which performers can purposefully adjust their intonation based on timbre and harmonic context. A continuation of this work would require refined experimental designs with a larger pool of subjects and more variation in timbre combinations to open up enquiries about specific instrumental pairs that only differ on the basis of timbre. A greater number of subjects allows for more reliable investigations into between-group differences such as instrument familiarity, as well as within-group differences stemming from tuning and rating styles. The length of the supervised training period at the start of future tuning experiments could be included as a between-subjects factor for familiarity, to see if it noticeably affects the recorded results. In addition, on top of examining how subjective intonation deviates from equal temperament (and/or other tuning systems), an in-depth examination of how subjective tuning affects the spectral shape of each timbre pair will help answer how much weight the harmonicity of an individual instrument has in comparison to the global spectral form of the composite timbre. Such methods could contribute to a clearer approach to understanding and overcoming the barriers between perception and performance in musical settings.

Appendix A

Experimental Stimuli

Table A.1 Description of experimental stimuli

Sample	Instrument	Note	Brightness	Playing Mechanism
1	Harpsichord (8’ + 4’ stop)	C4	High	Impulsive
2	Harpsichord (8’ + 4’ stop)	C#4	High	Impulsive
3	Harpsichord (8’ + 4’ stop)	E4	High	Impulsive
4	Harpsichord (8’ + 4’ stop)	F#4	High	Impulsive
5	Harpsichord (8’ + 4’ stop)	G4	High	Impulsive
6	Harpsichord (8’ + 4’ stop)	A4	High	Impulsive
7	Harpsichord (8’ + 4’ stop)	A#4	High	Impulsive
8	Harpsichord (8’ + 4’ stop)	C5	High	Impulsive
9	Harpsichord (8’ stop)	C4	High	Impulsive
10	Harpsichord (8’ stop)	C#4	High	Impulsive
11	Harpsichord (8’ stop)	E4	High	Impulsive
12	Harpsichord (8’ stop)	F#4	High	Impulsive
13	Harpsichord (8’ stop)	G4	High	Impulsive
14	Harpsichord (8’ stop)	A4	High	Impulsive
15	Harpsichord (8’ stop)	A#4	High	Impulsive
16	Harpsichord (8’ stop)	C5	High	Impulsive
17	Piano	C4	Medium	Impulsive
18	Piano	C#4	Medium	Impulsive
19	Piano	E4	Medium	Impulsive
20	Piano	F#4	Medium	Impulsive
21	Piano	G4	Medium	Impulsive
22	Piano	A4	Medium	Impulsive
23	Piano	A#4	Medium	Impulsive
24	Piano	C5	Medium	Impulsive
25	Clarinet	C4	Medium	Sustained
26	Clarinet	C#4	Medium	Sustained
27	Clarinet	E4	Medium	Sustained
28	Clarinet	F#4	Medium	Sustained
29	Clarinet	G4	Medium	Sustained
30	Clarinet	A4	Medium	Sustained
31	Clarinet	A#4	Medium	Sustained
32	Clarinet	C5	Medium	Sustained
33	Flute	C4	Low	Sustained
34	Flute	C#4	Low	Sustained
35	Flute	E4	Low	Sustained
36	Flute	F#4	Low	Sustained
37	Flute	G4	Low	Sustained
38	Flute	A4	Low	Sustained

Sample	Instrument	Note	Brightness	Playing Mechanism
39	Flute	A#4	Low	Sustained
40	Flute	C5	Low	Sustained
41	Trumpet	C4	High	Sustained
42	Trumpet	C#4	High	Sustained
43	Trumpet	E4	High	Sustained
44	Trumpet	F#4	High	Sustained
45	Trumpet	G4	High	Sustained
46	Trumpet	A4	High	Sustained
47	Trumpet	A#4	High	Sustained
48	Trumpet	C5	High	Sustained
49	Vibraphone	C4	Low	Impulsive
50	Vibraphone	C#4	Low	Impulsive
51	Vibraphone	E4	Low	Impulsive
52	Vibraphone	F#4	Low	Impulsive
53	Vibraphone	G4	Low	Impulsive
54	Vibraphone	A4	Low	Impulsive
55	Vibraphone	A#4	Low	Impulsive
56	Vibraphone	C5	Low	Impulsive

Appendix B

Experiment 1 Instructions

Timbre and Interval Tuning Study

- The goal of the experiment is to test the effects of timbre and sensory dissonance (auditory roughness) on tuning at various pitch intervals.

- The experiment is divided into 3 blocks. The first block consists of 8 practice trials, during which the experimenter will be there to answer any questions. The second and third blocks consist of 40 trials each.

- Each trial will display a screen that shows the pitch interval you are required to adjust. Your job is to tune the upper note of the given interval using small (3 cents) or big (10 cents) steps to get as close as you need to achieve the most acoustically in-tune sound as possible for that interval. After every adjustment, you can press the 'play' button to hear the result. Once you are satisfied with your tuning, you can press the 'done' button to move on to the next trial.

- Some of the trials will have different instruments as the upper and lower notes of the given interval, while some will have the same instrument on the upper and lower notes.

- When you have completed all the trials, the application will close. Please exit the sound booth with all of your belongings. You will be asked to fill out a short questionnaire regarding your musical background.

- Please ask the experimenter if you have any questions at this point.

Appendix C

Experiment 3 Instructions

Timbre and Interval Rating Study

- The goal of the experiment is to investigate the effects of instrumental timbre and musical consonance on the perception of auditory roughness (sensory dissonance), blend (perceptual fusion) and pleasantness at various pitch intervals.

    - **Auditory roughness** refers to the sensation of dissonance that occurs from detecting beating (amplitude fluctuations) between the partials of a sound event. Simply put, a rough sound may have a jarring, harsh, or raspy sonic texture.
    - **Blend**, or **perceptual fusion**, refers to the merging of two or more sound events into one unique sound event. A completely blended sound event will be wholly unified, whereas a completely unblended sound event will have independently audible parts. A partially blended sound event may fall somewhere between these two perceptual extremes.

- The experiment is divided into five blocks. The first block has 8 practice trials, during which the experimenter will be available to answer any questions. The next four blocks consist of 44 trials each.

- Each trial will display a screen with three continuous scales you along which you are required to make your rating. Your job is to rate the given dyads on their perceptual roughness, blend and pleasantness. Please focus on the auditory characteristics of the instrumental sounds rather than the musical significance of the intervals when making your ratings. You may use the 'replay' button to listen to the sound again for a total of three replays. Once you are satisfied with your ratings, press the 'done' button to move on to the next trial.

- Some of the trials will have different instruments as the upper and lower notes of the given interval, but others will have the same instrument on the upper and lower notes.

- When you have completed all of the trials, the application will close. Please exit the sound booth with all of your belongings. You will be asked to fill out a short questionnaire regarding your musical background at the end of the session.

- Please ask the experimenter if you have any questions at this point.

Bibliography

Audacity Team (2017). Audacity®: Free Audio Editor and Recorder [Computer application]. Version 2.1.2 retrieved January 12, 2016 from https://audacityteam.org/.

Bidelman, G. M., & Heinz, M. G. (2011). Auditory-nerve responses predict pitch attributes related to musical consonance-dissonance for normal and impaired hearing. *The Journal of the Acoustical Society of America, 130*(3), 1488–1502.

Bidelman, G. M., & Krishnan, A. (2009). Neural correlates of consonance, dissonance, and the hierarchy of musical pitch in the human brainstem. *The Journal of Neuroscience, 29*(42), 13165–13171.

Brues, A. M. (1927). The fusion of non-musical intervals. *The American Journal of Psychology, 38*(4), 624–638.

Costa, M., Ricci Bitti, P. E., & Bonfiglioli, L. (2000). Psychological connotations of harmonic musical intervals. *Psychology of Music, 28*, 4–22.

Cousineau, M., McDermott, J. H., & Peretz, I. (2012). The basis of musical consonance as revealed by congenital amusia. *Proceedings of the National Academy of Sciences of the United States of America, 109*(48), 19858–19863.

DeWitt, L. A., & Crowder, R. G. (1987). Tonal fusion of consonant musical intervals: The oomph in Stumpf. *Perception & Psychophysics, 41*(1), 73–84.

Di Stefano, N., & Bertolaso, M. (2014). Understanding musical consonance and dissonance: Epistemological considerations from a systemic perspective. *Systems, 2*, 566–575.

Geringer, J. M., & Witt, A. C. (1985). An investigation of tuning performance and perception of string instrumentalists. *Bulletin of the Council for Research in Music Education, 85*, 90–101.

Geringer, J. M., & Worthy, M. D. (1999). Effects of tone-quality changes on intonation and tone-quality ratings of high school and college instrumentalists. *Journal for Research in Mathematics Education, 47*(2), 135–149.

Hall, D. E., & Hess, J. T. (1984). Perception of musical interval tuning. *Music Perception: An Interdisciplinary Journal, 2*(2), 166–195.

Hasegawa, R. (2009). Gérard Grisey and the ‘nature’ of harmony. *Music Analysis, 28*/ii-iii*,* 349–371.

Hutchins, S., Roquet, C., & Peretz, I. (2012). The vocal generosity effect: How bad can your singing be? *Music Perception: An Interdisciplinary Journal, 30*(2), 147–159.

Hutchinson, W., & Knopoff, L. (1978). The acoustic component of western consonance. *Interface, 7,* 1–29.

ISO 389–8. (2004). *Acoustics – Reference zero for the calibration of audiometric equipment – Part 8: Reference equivalent threshold sound pressure levels for pure tones and circumaural earphones* (Tech. Rep.). Geneva, Switzerland: International Organization for Standardization.

Johnson-Laird, P. N., Kang, O. E., & Leong, Y. (2012). On musical dissonance. *Music Perception: An Interdisciplinary Journal, 30*(1), 19–35.

Kameoka, A., & Kuriyagawa, M. (1969a). Consonance theory part I: Consonance of dyads. *The Journal of the Acoustical Society of America, 45*(6), 1451–1459.

Kameoka, A., & Kuriyagawa, M. (1969b). Consonance theory part II: Consonace of complex tones and its calculation method. *The Journal of the Acoustical Society of America, 45*(6), 1460–1469.

Karrick, B. (1998). An examination of the intonation tendencies of wind instrumentalists based on their performance of selected harmonic intervals. *Journal of Research in Music Education, 46*(1), 112–127.

Kendall, R., & Carterette, E. (1993). Identification and blend of timbres as a basis for orchestration. *Contemporary Music Review*, *9*, 51–67.

Kopiez, R. (2003). Intonation of harmonic intervals: Adaptability of expert musicians to equal temperament and just intonation. *Music Perception, 20*(4), 383–410.

Krumhansl, C. (1991). Music psychology: Tonal structures in perception and memory. *Annual Review of Psychology, 42*, 277–303.

Lakatos, S. (2000). A common perceptual space for harmonic and percussive timbres. *Perception & Psychophysics*, *62*(7), 1426–1439.

Lembke, S. & McAdams, S. (2015). The role of spectral-envelope characteristics in perceptual blending of wind-instrument sounds. *Acta Acustica united with Acustica, 101,* 1039–1051.

Martin, F. N., & Champlin, C. A. (2000). Reconsidering the limits of normal hearing. *Journal of the American Academy of Audiology*, *11*(2), 64–66.

Mashinter, K. (2006). Calculating sensory dissonance: Some discrepancies arising from the models of Kameoka & Kuriyagawa, and Hutchinson & Knopoff. *Empirical Musicology Review, 1*(2), 65–84.

McAdams, S. (1984). The auditory image: A metaphor for musical and psychological research on auditory organization. *Advances in Psychology, 19,* 289–323.

McAdams, S., Winsberg, S., Donnadieu, S., De Soete, G., & Krimphoff, J. (1995). Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. *Psychological Research*, *58*(3), 177–192.

McDermott, J. H., Keebler, M. V., Micheyl, C., & Oxenham, A. J. (2010). Musical intervals and relative pitch: Frequency resolution, not interval resolution, is special. *The Journal of the Acoustical Society of America, 128*(4), 1943–1951.

Moore, B. C. J., Glasberg, B. R., & Peters, R. W. (1986). Thresholds for hearing mistuned partials as separate tones in harmonic complexes. *The Journal of the Acoustical Society of America, 80*, 479–483.

Moran, H., & Pratt, C. C. (1926). Variability of judgments on musical intervals. *Journal of Experimental Psychology, 9*(6), 492–499.

Platt, J. R., & Racine, R. J. (1985). Effect of frequency, timbre, experience, and feedback on musical tuning skills. *Perception & Psychophysics, 38*(6), 543–553.

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

Pressnitzer, D., McAdams, S., Winsberg, S., & Fineberg, J. (2000). Perception of musical tension for nontonal orchestral timbres and its relation to psychoacoustic roughness. *Perception and Psychophysics, 62*(1), 66–80.

Rakowski, A. (1990). Intonation variants of musical intervals in isolation and in musical contexts. *Psychology of Music, 18*, 60–72.

Sandell, G. J. (1995). Roles for spectral centroid and other factors in determining “blended” instrument pairings in orchestration. *Music Perception*, *13*, 209–246.

Schellenberg, G. (2002). Asymmetries in the discrimination of musical intervals: Going out-of-tune is more noticeable than going in-tune. *Music Perception, 19*(2), 223–248.

Sethares, W. A. (1993). Local consonance at the relationship between timbre and scale. *Journal of the Acoustical Society of America, 94*(3), 1218–1228.

Sethares, W. A. (1994). Adaptive tunings for musical scales. *Journal of the Acoustical Society of America, 96*(1), 10–18.

Sethares, W. A. (1998). *Tuning, Timbre, Spectrum, Scale.* London, England: Springer-Verlag.

Shields, R. & Kendall, R. (2004). The relation of timbre to dissonance and spectral fusion.

Proceedings of the 8th International Conference on Music Perception & Cognition, 596–599.

Smith, B. K. (1995). PsiExp: An environment for psychoacoustic experimentation using the IRCAM musical workstation. In *Society for Music Perception and Cognition Conference'95*. Berkeley, CA: University of California, Berkeley.

Swaffield, W. R. (1974). Effect of melodic parameters on ability to make fine-tuning responses in context. *Journal of Research in Music Education, 22*(4), 305–312.

Tenney, J. (1988). *A History of* '*consonance*' *and* '*dissonance.*' New York: Excelsior Music Publishing Company.

Terhardt, E. (1984). The concept of musical consonance: A link between music and psychoacoustics. *Music Perception: An Interdisciplinary Journal, 1*(3), 276–295.

Vos, J. (1982). The perception of pure and mistuned musical fifths and major thirds: Thresholds for discrimination, beats, and identification. *Perception & Psychophysics, 32*(4), 297–313.

Vurma, A., Raju, M., & Kuuda, A. (2010). Does timbre affect pitch? Estimations by musicians and non-musicians. *Psychology of Music, 39*(3), 291-306.

Wapnick, J. & Freeman, P. (1980). Effects of dark–bright timbral variation on the perception of flatness and sharpness. *Journal of Research in Music Education, 28*(3), 176–184.

Worthy, M. D. (2000). Effects of tone-quality conditions on perception and performance of pitch among selected wind instrumentalists. *Journal of Research in Music Education, 48*(3), 222–236.