Modelling Intonation and Interaction in Vocal Ensembles

Queen Mary

University of London

Jiajie Dai

School of Electronic Engineering and Computer Science

Queen Mary University of London

A thesis submitted for the degree of

Doctor of Philosophy

Mar. 2019

"If I cannot fly, let me sing."

By Stephen Sondheim

Acknowledgements

The first person I want to thank is my supervisor Prof. Simon Dixon. Prof. Dixon has been incredibly helpful for this work. I met a lot of issues during experiments, and Prof. Dixon has always given me the strongest support and helps me with great patience. He has been a great help with large conceptual issues as well as small experiment design. His detail-oriented personality has influenced me a lot. I highly appreciated his positive encouragement during the project. His paper *Intonation in Unaccompanied Singing* inspired me and gave me a lot of guidance. It makes me really enjoy my works and make a contribution to this area. I can never thank him enough for the help.

I'd like to say thank you to my second supervisors Dr. Dan Stowell and Dr. Matthias Mauch as well. This project would not have been what it is without Dan and Matthias' help since they gave me many suggestions and instruction. They assisted me greatly when I was confused about some statistical tests and mathematical modelling. I am very grateful to my academic assistant Dr. Marcus Pearce. I would also like to thank him for the helpful comments on my first journal paper.

Many thanks to all the participants that helped me during this project. I will never forget those days that we shared happiness and bitterness from this research. Many thanks to all the musical societies that helped me during this project, including the QMUL *A Capella* Society, QMUL Music Society, London Philharmonic Choir, the Hi-Fan Vocal Group.

Abstract

Voice is our native instrument and singing is the most universal form of music-making. As an important feature of singing, intonation accuracy has been investigated in previous studies, but the effect of interaction between singers has not been explored in detail. The aim of this research is to investigate interaction between singers in vocal ensembles, with a particular emphasis on how singers negotiate a joint reference pitch as the music unfolds over time. This thesis reports the results of three experiments which contribute to the scientific understanding of intonation.

The first experiment tested how singers respond to controlled stimuli containing time-varying pitches. It was found that time-varying stimuli are more difficult to imitate than constant pitches, as measured by absolute pitch error. The results indicate that pitch difference, transient duration, and stimulus type have a significant influence on pitch error, and the instability of the acoustic reference has a positive correlation with pitch error.

The second experiment measured pitch accuracy and interaction in unaccompanied unison and duet singing. The results confirm that interaction exists between vocal parts and influences the intonation accuracy. The results show that vocal part, singing condition (unison or duet), and listening condition (with or without a partner) have a significant effect on pitch accuracy, which leads to a linear mixed effect model describing the interaction by effect size and influencing factors.

In the third experiment, the effect of interaction on both intonation accuracy and the pitch trajectory was tested in four-part singing ensembles. The results show: singing without the bass part has less mean absolute pitch error than singing with all vocal parts; mean absolute melodic interval error increases when participants can hear the other parts; mean absolute harmonic interval error is higher in the one-way interaction condition than the two-way interaction condition; and the shape of note trajectories varies according to adjacent pitch, musical training and sex.

Statement of Originality

I, Jiajie Dai, confirm that the research included within this thesis is my own work or that where it has been carried out in collaboration with or supported by others, that this is duly acknowledged and my contribution indicated. Previously published material is also acknowledged herein.

I attest that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge break any UK law, infringe any third parties copyright or other Intellectual Property Right, or contain any confidential material.

I accept that the College has the right to use plagiarism detection software to check the electronic version of the thesis.

I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university.

Jiajie Dai

Contents

List of Figures

10

List of Tables

List of Variables and Acronyms

The following variables and acronyms are used within the body of the thesis.

$\bar{p}_i$ Median of the observed pitch

$\hat{d}$ Modelled length of transient part

$\hat{p_D}$ Modelled pitch difference

$\hat{p_m}$ Modelled main pitch

c_i Reference pitch

d Length of transient part

e^h Harmonic interval error

e^i Interval error

e^m Melodic interval error

e^p Pitch error

f_0 Fundamental frequency

f_{ref} Reference frequency

i Index of note

p Pitch in semitones

$p(t)$ Pitch trajectory

p^s Score pitch

p_D Pitch difference

p_m Main pitch

$p_r(t)$ Pitch of response note

$p_s(t)$ Pitch of the stimulus

p$_{ref}$ Reference pitch

ANOVA Analysis of Variance

LMER Linear mixed-effects regression

MAHIE Mean absolute harmonic interval error

MAMIE Mean absolute melodic interval error

MAPE Mean absolute pitch error

MIDI Musical Instrument Digital Interface

MIE   Melodic interval error

MPE   Mean pitch error

OMR Optical music recognition

SATB Choirs comprises four musical parts: soprano, alto, tenor and bass

SD   Standard deviation

TRD   Tonal reference deviation

Chapter 1

Introduction

This thesis is concerned with the intonation and interaction between singers in vocal ensembles. In this chapter we explain the motivations and aim of our work (Section 1.1 and 1.2). Also the structure of the thesis is provided (Section 1.3) along with the main contributions of this work. Finally, Section 1.4 concludes the chapter with a list of our own publications relating to the thesis.

1.1 Motivation

Voice is our original instrument (La Barbara, 2002), even from prehistoric times (Mithen, 2007), and it is one of the defining features of humanity (Welch, 2005). This instrument communicates emotion, expressing joy and sadness, hope and despair. Singing as one use of the voice has a wonderful and unique expression which makes it the most complex form of all the performance arts (Miller, 1996), yet the factors that determine singing proficiency are still poorly understood.

Intonation's extreme importance in Western music arises from the fact that it relates to both melody and harmony. People have wondered about how to get better pitch accuracy for a long time. There are abundant publications teaching people how to sing better, while fine singers seldom analyse the intonation variations with computing and modelling. People objectively ascribe their inaccuracy in pitch to talent and training, but few of them address the problems by scientifically investigating singing ensembles. In recent years, the problem of intonation has gained considerable research interest due to the popularisation of vocal training and performance, as well as the development of algorithms and research in the area. Although intonation accuracy in singing has been studied (e.g. Pfordresher and Brown, 2007; Mauch et al., 2014), there are many aspects such as interaction and

pitch trajectory within the notes which need further investigation.

Tones produced by the singing voice are unlike those of most other pitched instruments, in that their time-varying pitch trajectories within the notes are quite irregular (e.g. Gerhard, 2005; Mauch et al., 2014; Dai et al., 2015). Analysis of the vocal imitation of pitch trajectories is one of the approaches which helps us get a better understanding of the topic. A vocal imitation task involves both the perception and production of such tones and reflects a common paradigm for learning music. Although singing in tune is a primary element of singing performance, vocal imitation of time-varying pitch stimuli has not been researched. This motivates Chapter 3 on vocal imitations of pitch trajectories, in order to distinguish and quantify the interaction of participants when tuning with an artificial stimulus.

Singers without instrumental accompaniment can react to deviations in intonation by other singers and will do so to maintain a harmonically pleasing performance. Yet little is known on how this interaction happens, even though it plays an important role in ensemble performance. Terasawa (2004) claimed that the intonation accuracy of choral members was influenced by the progression of chord roots. Brandler and Peynircioglu (2015) observed that participants learned new pieces of music more efficiently when learning it individually than with companions. Mürbe et al. (2002) observed that singers' intonation accuracy is reduced in the absence of auditory feedback. Although many publications give guidelines to keep singers in tune by training them as excellent soloists (e.g. Bohrer, 2002; Alldahl, 2008), the interaction between singers as it unfolds in real-time has not been fully researched, which motivates Chapter 4 to improve the scientific understanding of the interaction between dual singers.

In Western music, one common configuration for singing ensembles and choirs comprises four musical voices or parts: soprano, alto, tenor and bass (SATB); so we chose the SATB ensemble as the research target for the experiment. SATB ensembles are ideal to explore the complex interactions in a group of singers. The motivation of Chapter 5 is to test the influence of the various vocal parts and how the singers interact with each other, especially how hearing other singers influences the performance of each vocal part. These effects are tested in terms of their effect on intonation.

Besides the intonation accuracy, it is interesting to know whether there are patterns or regularities in the pitch trajectories of individual notes. Chapter 6 is a further extension of Chapter 5 which studies pitch variation within vocal notes and ascertains what factors influence the various parts of a note. Chapter 6 aims to find common trends in the note

trajectories, with differences due to context and experimental conditions.

1.2 Aim and Research Questions

The main aim of the project is an improved understanding of the complex interactions that occur when people sing together, and more specifically the interactions described in the three stages: the dynamics of one singer's intonation in response to a synthetic stimulus; the interaction between two singers; and the interaction between four singers. A particular outcome will be analysing and modelling of two-way singer interaction. I expect that these outcomes in turn will have impact in society by allowing conductors to understand the most effective way of improving intonation for the individual singer, and in their ensembles and choirs.

The general research aim of this project is to explain interaction and intonation in vocal ensembles. To reach this target, we separate our project into several specific research questions:

- How does the intonation change when a singer responds to a pitch-varying stimulus?

- What do singers do when they sing together?

- What role does interaction play in SATB singing?

- Are there any patterns in the pitch trajectories?

This thesis attempts to achieves these targets, by (1) conducting experiments with groups of singers, (2) analysing and labelling the recordings obtained with participants' information, score information and pitch accuracy, and by (3) developing mathematical models to explain the effects observed.

1.3 Thesis Structure

Chapter 1: Introduction

In this chapter we identify motives for each experiment and define the aim and research questions. The relevant publications and the thesis's main contributions are included.

Chapter 2: Background and Previous Work

Introduces the main bodies of existing research which we will build upon. It begins by presenting an overview of related work on the area and then surveys relevant research

topics including intonation accuracy, accompaniment, pitch drift and relative tools. There is a special focus on the intonation metrics that will be used in the following chapters. Then a preliminary project is presented as the preparation of the research aim.

Chapter 3: Imitation Study

Investigates the pitch trajectories of vocal imitations by individual singers. To achieve the aim, five stimulus types were designed for participants to imitate simultaneously or alternately with the stimulus. The data was collected by a customised online system. After producing the annotated data, the significant factors were tested and the accuracy for different stimuli were compared. In particular, parameters of the responses were extracted by a forced fit to a model of the stimulus type, in order to describe the observed pitch trajectories.

Chapter 4: Duet Interaction Study

Investigates singing interaction by analysis of the factors influencing pitch accuracy of unaccompanied duet singers. The experiment investigates singing conditions and listening conditions which have a significant influence on intonation. Models were presented to describe the factor of interaction. In particular, singing with the same vocal part is more accurate than singing with a different vocal part; singing solo has less pitch error than singing with a partner; other factors influence the pitch accuracy, including: score pitch, score harmonic interval, score melodic interval, musical background, vocal part and individual differences.

Chapter 5: SATB Study

This chapter investigates interactions in four-part (SATB) singing from the point of view of pitch accuracy (intonation) by comparing intonation accuracy of individual singers and collaborative ensembles. A novel experiment tested the intonation accuracy of five groups of singers in a series of test and listening conditions. The results confirm that interaction exists between singers and which factors influence their intonation, and that intonation accuracy depends on which other singers each individual singer can hear. More specifically: Singing without the bass part has less mean absolute pitch error than singing with all vocal parts; mean absolute melodic interval error increases when participants can hear the other parts.

Chapter 6: Intonation Trajectories and Patterns of Vocal Notes

This chapter presents a general pattern of vocal notes which possess transient components in the beginning and end of a note, although the shape of the pattern may vary according

to the individual performer, previous pitch, next pitch, vocal part and sex. According to the analysis of over 35000 individual notes, a general shape of vocal notes was found which contains transient components at the beginning and end of each note. A general expansion of harmonic intervals was observed: about 8 cents pitch difference is observed between adjacent vocal parts, with sopranos singing sharper and male singers flatter than the target pitch.

Chapter 7: Conclusions and Future Perspectives

This chapter provides a summary of the achievements of this thesis and future perspectives on further study and potential applications of this research.

1.4 Associated Publications

This thesis covers work on intonation and interaction analysis which was carried out by the author between December 2014 and November 2018 at Queen Mary University of London under the supervision of Simon Dixon. The majority of the work presented in this thesis has been presented in international peer-reviewed conferences.

Journal Paper

  (i) Dai, J. and Dixon, S. (2019b). Singing Together: Pitch Accuracy and Interaction in Unaccompanied Duet Singing. *The Journal of the Acoustical Society of America*, 145(2):663–675

Peer-Reviewed Conference Papers

  (ii) Dai, J., Mauch, M., and Dixon, S. (2015). Analysis of Intonation Trajectories in Solo Singing. In *16th International Society for Music Information Retrieval Conference*, pages 420–426

  (iii) Dai, J. and Dixon, S. (2016). Analysis of Vocal Imitations of Pitch Trajectories. In *17th International Society for Music Information Retrieval Conference*, pages 87–93

  (iv) Dai, J. and Dixon, S. (2017). Analysis of Interactive Intonation in Unaccompanied SATB Ensembles. In *18th International Society for Music Information Retrieval Conference*, pages 599–605

  (v) Dai, J. and Dixon, S. (2019c). Understanding Intonation Trajectories and Patterns of Vocal Notes. In *25th International Conference on MultiMedia Modeling*, pages 243–253

Other Publications

(vi) Mauch, M., Cannam, C., Bittner, R., Fazekas, G., Salamon, J., Bello, J., Dai, J., and Dixon, S. (2015). Tony: a Tool for Efficient Computer-aided Melody Note Transcription. In *the First International Conference on Technologies for Music Notation and Representation (TENOR)*

Under Review

(vii) Dai, J. and Dixon, S. (2019a). Intonation Trajectories in Unaccompanied SATB Singing. .

Publication (ii) is based on the author's master's thesis; although the experiment design and data collection were finished in master period, the analysis and modelling were performed in the first year of the PhD. It should be noted that for paper (vi) the author contributed to the collection of the test data.

1.5 Conclusions

After considering the motivations for intonation and interaction analysis, this chapter has stated the goal of this thesis: to understand and model the interaction in terms of intonation in singing ensembles. Then the chapter has laid out the structure of the thesis and indicated the four chapters that contain its main contributions: the imitation study by individual singers (Chapter 3), the duet study which investigates the effects of listening conditions and singing conditions (Chapter 4), and the SATB study which extends a standard musical situation (Chapter 5) and explores the transient parts of notes (Chapter 6). The thesis contributes novel experimental designs and public data sets. The literature review in the following chapter brings together all the information needed to understand how the research goal has motivated the design of experiments.

Chapter 2

Background and Previous Work

This chapter covers a basic understanding of musical concepts such as intonation, and related work in pitch accuracy and interaction, which are prerequisites to the development of the new investigations in this thesis. Firstly, some terms from music theory will be introduced, which will be used throughout the thesis (Section 2.1). The cores of this chapter include a description of the components and operation of the human vocal system (Section 2.2), a detailed review of intonation and metrics (Section 2.3), followed by a discussion of pitch drift (Section 2.4). The software and tools in this area, especially the tools used in this project, are introduced in Section 2.5. Finally, a preliminary project is described which informed this current work (Section 2.6).

2.1 Relevant Music Terminology

Music is an art form and cultural activity whose medium is sound organised in time. The common elements of music are pitch (melody and harmony), rhythm (tempo, meter, and articulation), dynamics (loudness and softness), and the sonic qualities of timbre and texture (tone), and this thesis is about intonation and interaction of vocal ensembles. The following sections will introduce relevant concepts.

2.1.1 Pitch

Pitch is a psychoacoustic percept that does not map simply onto physical properties of sound (Walker et al., 2011). It is one of the most important concepts in tonal music, and harmony builds upon the human ability to perceive pitch (Mauch, 2010). The American National Standards Institute (ANSI) defines pitch as the auditory attribute of sound according to which sounds can be ordered on a scale from low to high (American National

Standards Institute, 1973).

The fundamental frequency (F_0), often referred to simply as the fundamental, it is normally the lowest frequency of a periodic waveform. For a periodic waveform with period T in seconds, the fundamental frequency F_0 in Hz is given by $F_0 = \frac{1}{T}$. For pitched musical sounds, which are almost periodic, the perceived pitch corresponds quite closely to the fundamental frequency; and they are often used interchangeably. There are some non-linear perceptual effects that alter perceived pitch such as timbre (and level if the stimulus is a sinewave) (Howard and Angus, 2017).

For voiced sounds in speech or singing, the fundamental depends on the frequency of the vocal fold that is opening and closing. Because at any particular point in speech the folds are executing a roughly periodic mechanical pattern, the sound wave emitted is also periodic.

Usually, the range of human hearing is from 20 Hz to 20 kHz, depending on age and individual. The voiced speech of a typical adult male will have a fundamental frequency from 85 to 180 Hz, and that of a typical adult female from 165 to 255 Hz (Titze and Martin, 1998; Baken and Orlikoff, 2000), however, the fundamental frequency of the singing voice has a wider range. Although fundamental frequency in Hertz and pitch in semitones are both common in music research, pitch in semitones was chosen for this thesis because it corresponds more closely to perception (Pfordresher, 2012).

The range of pitches generally used in music covers a little more than seven octaves which is the general range covered by a concert grand piano. The fundamental frequency ranges of instruments and the human voice are different (shown Figure 2.1). The notes produced by pitched instruments are typically stable in terms of their fundamental frequency, and they are heard as notes of a definite pitch. However, the pitch of the human voice is not as stable as the pitch of instruments, motivating the studies in Chapters 4 and 6. (Literature on the human vocal system is reviewed in Section 2.2.)

In order for different instruments and vocal ensembles to work together, tuning is necessary. For most instruments, the players should ideally tune to the same reference pitch (note A4 at 440Hz for example) heard on another instrument, then all the instruments sound in tune with each other. For singing ensembles, they have to rely on complex factors for tuning (more discussion can be found in section 2.2).

In empirical research, fundamental frequency is usually measured in Hertz using computer software and then converted to pitch in semitones with Equation 2.1, where p is the pitch in semitone, F_0 is the measured fundamental frequency and f_{ref} is the reference frequency.
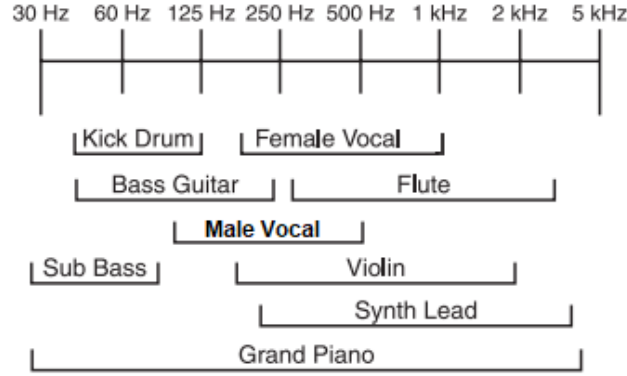
Figure 2.1: Fundamental frequency ranges of various instruments (adapted from Hewitt (2008)).

In the equation, p is not the pitch as in the note name (e.g. A4) but is instead the number of semitones above the reference pitch (and the reference pitch has its own particular value).

$$p = 12 \times \log_2(\frac{F_0}{f_{\mathrm{ref}}}) + p_{\mathrm{ref}} \qquad (2.1)$$

Western music often uses a collection of pitches that divide the octave into 12 equal parts (semitones), each of which can be further divided into 100 parts (cents). This thesis measures pitch with the standard A4 piano key pitch as the reference pitch, where $f_{\mathrm{ref}} = 440$Hz. The pitch scale used corresponds with the *MIDI* (Musical Instrument Digital Interface) specification, where $p_{\mathrm{ref}} = 69$ for A4. Pitch and fundamental frequency are related logarithmically, meaning that linear changes in pitch like going up an octave are multiplicative in frequency, an octave specifically is doubling the fundamental frequency. For instance, from 400 to 800 Hz spans an octave, and so is from 800 Hz to 1600 Hz, because each is a multiplication by a factor of two. Although just noticeable difference for pitch is about 5 cents (Loeffler, 2006) and some of the effect sizes in this thesis are under this value, the accumulation of multiple factors may lead to a noticeable difference. Reporting the effect size under 5 cents still has statistical meaning.

Absolute pitch and relative pitch are two important concepts in this thesis. Absolute pitch, widely referred to as perfect pitch, is a rare auditory phenomenon characterised by the ability of a person to identify or re-create a given musical note without the benefit of a reference tone (Takeuchi and Hulse, 1993). Relative pitch is the ability of a person to identify or re-create a given musical note by comparing it to a reference note and identifying the interval between those two notes.

From 1 to 5 people per 10,000 have absolute pitch, according to estimates (Brown et al.,

2003). Perfect pitch occurs in musicians at higher rates, from less than 1% up to 11%, according to some studies (Baharloo et al., 2000). It runs in families, suggesting a genetic link, and occurs most often in people who had musical training before age 6. More of the population have relative pitch, and especially, musicians learn relative pitch as part of their training.

2.1.2 MIDI Notation

A musical score can be stored as a data type in many different ways, however, a common music representation protocol is the Musical Instruments Digital Interface (MIDI) protocol. Using the MIDI protocol, the specific pitch, onset, offset, and intensity of a note can be stored, along with additional parameters such as instrument type, key, and tempo.

The MIDI standard for pitch starts from a value of 0 which is 8.1758 Hz and extends to a value of 127 which is G9=12543 Hz. The conversion formula from frequency to MIDI is shown in Equation 2.1 with $f_{ref} = 440$ and $p_{ref} = 69$. For integer values of pitch the scale coincides with the MIDI standard. Note that pitch is not constrained to integer values in this representation.

MIDI notation is limited compared with original music recordings, or representations of musical notation, but the accessibility and simplicity make it appropriate for this project. Besides MIDI notation, there are many protocols used for music computing, such as MusicXML or Lilypond. Automatic transcription systems usually convert the input recording into a MIDI file or a MIDI-like representation.

All the results are represented using the MIDI standard in this thesis. Figure 2.2 is an example of a piano-roll representation of the first piece in Chapter 5 (a Bach chorale "Oh Thou, of God the Father"). Ideally, the score pitch should look like Figure 2.2 but the actual observation is different due to the participants.

2.1.3 Intervals

A musical interval is the difference between two pitches (Prout, 2011) and is proportional to the logarithm of the ratio of the fundamental frequencies of the two pitches. Based on Equation 2.1, the interval from a pitch p_1 to the pitch p_2 is defined as:

$$i = p_2 - p_1 \qquad (2.2)$$

However, it is difficult to keep the interval accuracy. There is a phenomenon called *compression* (Pfordresher et al., 2010) which means people sing melodic intervals smaller than

26
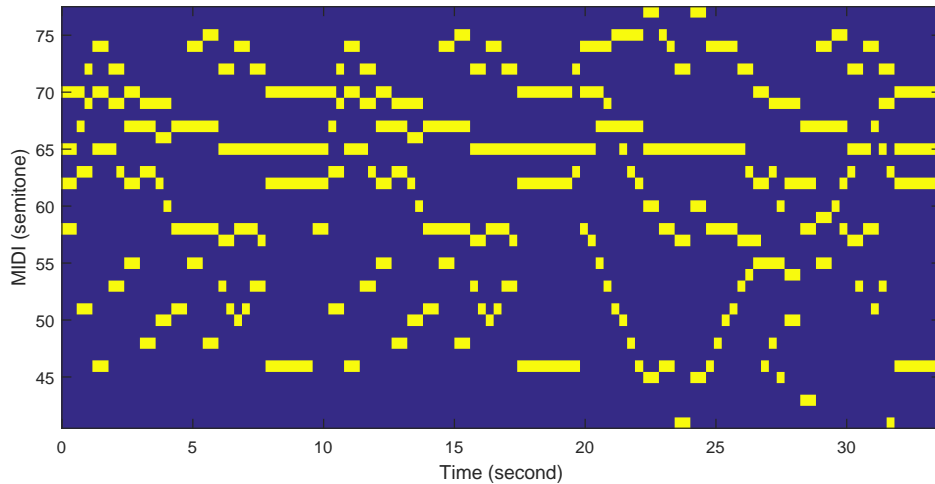
Figure 2.2: The piano-roll representation of a SATB piece

the correct interval. For example, consider a singer in the experiment who sang F4 at 364.4 Hz and C5 at 510.34 Hz for two adjacent notes. There should be a difference of 7 semitones between these two notes. But the observed difference is 5.83 semitones (using Equation 2.1 and 2.2) which is smaller than the target interval (7 semitones).

There are four intervals which are called perfect intervals, and are found in both major and minor scales. Perfect intervals include the unison (the same tone repeated), fourth (five semitones), fifth (seven semitones) and octave (twelve semitones).

2.1.4 Tuning Systems

The term *just intonation* describes the tuning of musical intervals so that their frequencies are related by small whole number ratios. Just intonation, also called pure intonation, results in harmonic intervals that do not beat (i.e. they sound more consonant), and melodic intervals derived from such an arrangement, result in more than one size of whole tone (Lindley, 2001). There is more than one way to get whole-number ratios (e.g. Pythagorean).

Justly in tune intervals have a unique quality of smoothness or purity, at least when used with many standard musical timbres. In the mathematical aspect, harmony is a combination of individual sounds where the frequencies of notes are related by ratios of small whole numbers. For example, the fundamental frequency of A4 is 440 Hz and of A5 is 880 Hz, the ratio of their frequencies is 1:2 and the relationship between them is called an octave in musical terms. For example, one of the important ratios is 3:2 (about 702 cents), which is called the perfect fifth.

Singers in a capella choirs appear to prefer to sing in just intonation which is based on

27

the use of integer ratios to derive the frequency ratios between the main musical intervals that make up a scale - in just intonation these are the octave (2:1), the perfect fifth (3:2) and the major third (5:4) (Howard, 2007a).

One important issue with using just intonation is accumulated drift. The accumulation of the difference between just intonation and twelve-tone equal temperament may lead to pitch drift. For example, perfect 5th up, octave down, perfect 5th, perfect 5th up, octave down, perfect 5th up, major 3rd down, which gives the following, using pure ratios: $(3/2) \times (1/2) \times (3/2) \times (3/2) \times (1/2) \times (3/2) \times (4/5) = 81/80$, although you should end up on the same pitch you started on. If the sequence was repeated, the drift would accumulate. Another issue with just intonation is its complexity. To solve these issues, from the 18th century, most instruments use a tuning system call twelve-tone equal temperament where each pitch has a fixed frequency and their ratios are roughly equal to those of their pure counterparts.

2.1.5 Duration and Tempo

Duration is the length of time a pitch, or tone, is sounded (Benward, 2014). There are two types of duration in this project: score duration and observed duration. The score duration is the theoretical note duration calculated using the tempo specified in the score (or given by a metronome). The observed duration is the actual length of the note from onset to offset that participants produced.

Tempo is the speed or pace of the beat of a given piece. In classical music, the tempo is typically indicated with an instruction at the start of a piece, either using conventional Italian terms or a metronome marking measured in beats per minute (bpm). For example, the 1st piece in Chapter 4, is 120 bpm which means the score duration for a quarter note, in this case the beat level, is $60/120 = 0.5$ second.

2.2 The Human Voice

The human voice is the most widely used musical instrument, and yet it is notoriously difficult to control (Pfordresher et al., 2010). Moreover, the human voice is the oldest musical instrument and can be traced back to prehistoric societies (Brown, 1991). Understanding the mechanism of human voice production is useful for forming the hypotheses of each experiment and supporting reasonable discussion of the observations in the results.
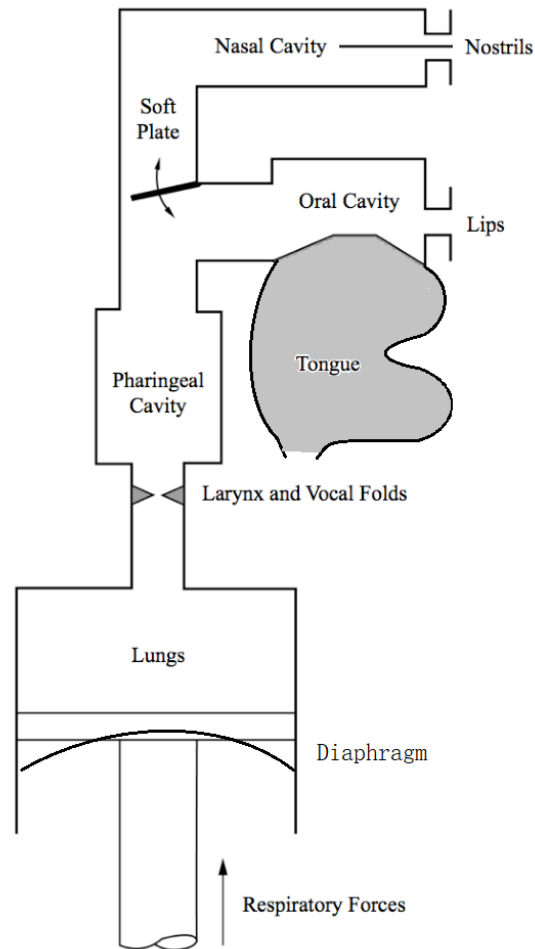
Figure 2.3: Model of the voice production apparatus, adapted from Clark and Yallop (1995).

2.2.1 The vocal production system

The human voice system can be thought of as a mechanical instrument. However, tones produced by the singing voice are unlike those of most other pitched instruments, in that their time-varying pitch trajectories are quite irregular (Gerhard, 2005; Mauch et al., 2014; Dai et al., 2015). Consider the voice organ as a generator of voiced sounds. Functionally the organ has three major units: a power supply (the lungs), an oscillator (the vocal folds) and a resonator (the vocal tract) (Sundberg, 1977). Each unit is composed of different parts of the body and has specific roles in voice production.

The lung is a spongy structure suspended within the rib cage. The small cavities in these spongy structures create a resonance, which is terminated by the vocal folds. The vocal folds are constituted by muscles shaped as folds and their covering is a mucous membrane. The combination of the pharynx and the mouth is referred to as the vocal tract. The nasal cavity is split up into two halves in its anterior parts (Figure 2.3).

The voice organ allows us to generate a great variety of voice sounds. Some of this family of sounds are speech sounds when the sounds are arranged in adequate sequences. In singing, there are both speech sounds and other types of sounds. The notes or tones can be regarded as a modified speech sounds. The differences between speech and singing are in the use of pitch range and vowel duration. Pitch or intonation is the main focus of this thesis. Since all the experiments asked participants to sing /ta/ or /a/ rather than the lyrics, the influence of vowels is not included. Although the syllable may influence the consonance, the factor of syllable does not show a significant effect on the pitch trajectory; more discussion can be found in Section 6.3.

2.2.2 Transient Parts

Various studies on singing have observed that vocal tones do not have a stable pitch (e.g. Gerhard, 2005; Mauch et al., 2014; Dai et al., 2015), but instead tend to fluctuate around a central pitch and exhibit transitions at their beginnings and ends.

The time-varying pitch of a vocal tone can be conceptualised as consisting of initial and final *transient* segments (where the statistical properties of the signal are rapidly changing and thus unpredictable) surrounding a longer *steady-state* section. This is similar to how the time-domain audio waveform is often modelled. For singing, such a segmentation is often difficult to perform, as the pitch signal rarely reaches a steady state.

At the beginning of a tone, a pitch glide is often observed as the singer adjusts the vocal folds from their previous state (the previous pitch or a relaxed state). Then the pitch is adjusted as the singer uses perceptual feedback to correct for any discrepancy between the auditory feedback and the intended note (Zarate and Zatorre, 2008). Possibly at the same time, vibrato may be applied, which is an oscillation around the central pitch, which is close to sinusoidal for skilled singers, but asymmetric for unskilled singers (Gerhard, 2005; Sundberg, 1994). Finally, they may not sustain the pitch at the end of the tone, and the pitch often moves in the direction of the following note, or downward (toward a relaxed vocal fold state) if there is no immediately following note (Xu and Sun, 2000).

2.2.3 Vibrato

Vibrato is a musical effect consisting of a regular, pulsating change of pitch. It is used to add expression to vocal and instrumental music. In singing, it can occur spontaneously through variations in the larynx. A good vibrato in music is a periodic pulsation, generally involving pitch, intensity, and timbre, which produces a pleasing versatile mellowness and

richness of tone (Seashore, 1931). Professional (particularly opera) singers tend to produce vibrato: a periodic modulation of F_0 , which is not normally used in speech (Sundberg, 1987).

All human voices can produce vibrato. In opera, as opposed to pop, vibrato begins at the starting of the note and continues to the end of the note with slight variations in width during the note. It has been shown that trained singers are able to elicit control over both vibrato rate and depth (Dromey et al., 2003; King and Horii, 1993).

There are three different voice vibrato processes that occur in different parts of the vocal tract: 1) The vocalis muscle vibrates at a frequency of 6.5 to 8 Hz.; 2) The diaphragm vibrates at a frequency below 5 Hz; 3) A combination of the two, resulting in a vibrato whose frequency is between 5 and 6.5 Hz. In performance, singers tend to use the combination (Fischer, 1993).

2.2.4 Factors Affecting Intonation Accuracy

Mürbe et al. (2002) showed how singers' intonation accuracy is reduced by diminished auditory feedback; in their experiment, auditory feedback was masked by noise. When singers cannot hear themselves, they have to rely on kinesthetic feedback circuits, which are less useful than auditory feedback for informing intonation. Likewise even in musical situations where the accompanying sound provides the tonal reference, singers make greater pitch errors when singing with accompaniment (Pfordresher and Brown, 2007), and particularly when the accompanying pitch content varies over the duration of a note (see Chapter 3). Thus vocal accompaniment is more difficult to sing with than instrumental accompaniment, because singers are relying on unstable reference pitches from other vocal parts (Liimola, 2000, p. 151). Although singing in unison with a partner may not increase pitch accuracy, it may give singers more confidence than singing solo (Heath and Gonzalez, 1995). However, Brandler and Peynircioglu (2015) observed that participants learned new pieces of music more successfully when in an individual learning environment than in a collaborative one.

Previous studies have investigated differences between solo and unison singing, although not all studies obtained significant results. For example, Green (1994) claimed that children singing unison, as opposed to individually, had significantly better vocal accuracy, while Cooper (1995) was unable to show a significant difference. Others observed that children sing more accurately individually than in a group (e.g. Clayton, 1986; Goetze, 1985, 1989). Besides the singing conditions, age, sex, training and number of attempts

were reported as significant factors for children's singing accuracy (e.g. Nichols and Wang, 2016; Nichols, 2016).

Except for the 0.01% of the population who have absolute pitch, the ability to identify or reproduce any given pitch on demand (Takeuchi and Hulse, 1993; Bohrer, 2002), most people rely on a reference pitch for tuning. An initial reference will be forgotten over time (Long, 1977; Mauch et al., 2014), so singers must constantly update their frame of reference as they sing, based on what they have recently heard, both their own voice and any accompaniment.

Much evidence shows that singers are influenced by other choral members in terms of pitch accuracy (e.g. Howard, 2003; Terasawa, 2004) and various approaches have been proposed to keep singers in tune by focusing on relative pitches, tone memories and muscle memories (e.g. Bohrer, 2002; Alldahl, 2008).

2.2.5 Individual Factors

For an individual singer, singing is a complicated task involving both perception and production. Factors related to production such as muscle strength and control can be improved by training and practice, while the perceptual factors involve many cognitive components with distinct brain substrates (Stewart et al., 2006).

Singers who exhibit much greater than average pitch errors are classified as *poor singers*, a phenomenon that has been the focus of several studies (Pfordresher and Brown, 2007; Dalla Bella et al., 2007; Berkowska and Dalla Bella, 2009; Pfordresher et al., 2010). This thesis intentionally chose non-poor singers as our research target. "Poor singers" are defined as those who have a deficit in the use of pitch during singing (Welch, 1979; Pfordresher and Brown, 2007), and are thus unable to perform the experimental tasks.

Participants whose pitch imitations had on average at least one semitone absolute error were categorised as poor singers. The data of poor-pitch singers (mean absolute pitch error larger than one semitone) is excluded in this thesis, apart from one singer who occasionally sang one octave higher than the target pitch.

For poor pitch singing, evidence points to a deficiency in pitch imitation accuracy as the main cause (Pfordresher and Mantell, 2014), although there are several types of singing deficiency and they vary by age and training (e.g. Demorest et al., 2015).

Vocal training is an important factor for enhancing the singing voice and making the singer's voice different from that of an untrained person (Mendes et al., 2003). Individual factors such as age and sex influence pitch accuracy (Welch et al., 1997). Musical training

and experience also have some influence on singing ability; Mauch et al. (2014) found that self-rated singing ability and choir experience, but not general musical background, correlated significantly with intonation accuracy.

To allow us to test for the effects of training, participants completed a questionnaire containing 34 questions from the Goldsmiths Musical Sophistication Index (Müllensiefen et al., 2011, 2014) which can be grouped into 4 main factors for analysis: active engagement, perceptual abilities, musical training and singing ability (9, 9, 7 and 7 questions respectively). A combined score of music background was calculated too (2 questions are not counted as main factors). The full marks for each factor are 63, 63, 49 and 49. The full questionnaire is in Appendix C. The score agreement can be found at the Musical Sophistication Index website (`https://www.gold.ac.uk/music-mind-brain/gold-msi/download/`).

2.3 Intonation Accuracy

Intonation is commonly regarded as an important aspect of music performance. Already in the 1930s, Seashore measured fundamental frequency (F_0) in recordings of renowned singers to reveal considerable departures from equally tempered tuning (Sundberg et al., 2013).

Intonation, defined as "accuracy of pitch in playing or singing" (Swannell, 1992), or "the act of singing or playing in tune" (Kennedy, 1980), is one of the main priorities in choir rehearsals (Ganschow, 2013) and in choral practice manuals (e.g. Crowther, 2003). Good intonation involves the adjustment of pitch to maximise the consonance of simultaneous notes, but it also has a temporal aspect, particularly in the absence of instrumental accompaniment, where the initial tonal reference can be forgotten over time (Mauch et al., 2014).

The definitions of intonation given above imply the existence of a reference pitch, which could be provided by accompanying instruments or could exist solely in the singer's memory. This latter case allows for the reference to change over time, and thus explain the phenomenon of drift.

2.3.1 Pitch Error

Assuming that a reference pitch has been given, *pitch error* can be defined as the difference between observed pitch and score pitch (Mauch et al., 2014):

$$e_i^p = p_i - p_i^s \qquad (2.3)$$
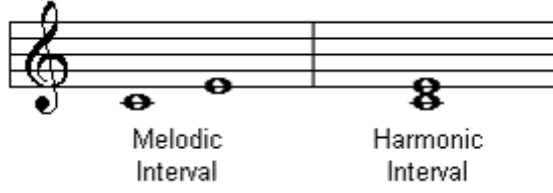
Figure 2.4: A melodic interval and harmonic interval of a major third (four semitones).

where p_i is the median of the observed pitch trajectory of note i (calculated over the duration of an individual note), and p_i^s is the score pitch of note i.

To evaluate the pitch accuracy of a sung part, the *mean absolute pitch error* (MAPE) is used. For a group of M notes with pitch errors $e_1^p, \ldots, e_M^p$, the MAPE is defined as:

$$\text{MAPE} = \frac{1}{M} \sum_{i=1}^{M} |e_i^p| \tag{2.4}$$

2.3.2 Interval Error

The interval error for the jth interval can be defined between a sung interval i_j and the expected nominal interval $i_n{}^s$ (given by the musical score) as:

$$e^i = i_j - i_n{}^s \tag{2.5}$$

The term *harmony* is used to describe the combination of concurrent pitches and the evolution of these note combinations over time (Benetos, 2012). A melodic interval refers to the pitch relationship between two consecutive notes while a melody refers to a series of notes arranged in a musically meaningful succession (Schoenberg, 1978). A harmonic interval occurs when two notes are played at the same time. In a *melodic interval*, the two notes are sounded in succession; while in a *harmonic interval*, both notes are played simultaneously (Figure 2.4).

The melodic interval error is calculated as the difference between the observed and score intervals:

$$e_i^m = (\bar{p}_{i+1} - \bar{p}_i) - (p_{i+1}^s - p_i^s) \tag{2.6}$$

where p_i^s and p_{i+1}^s are the score pitches of two sequenced notes, and $\bar{p}_i$ and $\bar{p}_{i+1}$ are their observed median pitches. Similarly, harmonic interval error is defined as:

$$e_{i,A,j,B}^h = (\bar{p}_{i,A} - \bar{p}_{j,B}) - (p_{i,A}^s - p_{j,B}^s) \tag{2.7}$$

where $p_{i,A}^s$ and $p_{j,B}^s$ are the score pitches of two simultaneous notes from singers A and B respectively, and $\bar{p}_{i,A}$ and $\bar{p}_{j,B}$ are their observed median pitches.

Harmonic intervals were evaluated for all pairs of notes which overlap in time. If one singer sings two notes while the second singer holds one note in the same time period, two harmonic intervals are observed. Thus indices i and j in Equation 2.7 are not assumed to be equal.

The *mean absolute melodic interval error* (MAMIE) for M intervals is calculated as follows:

$$\text{MAMIE} = \frac{1}{M} \sum_{i=1}^{M} |e_i^m|. \tag{2.8}$$

The *mean absolute harmonic interval error* (MAHIE) is calculated similarly (simplifying the notation and assuming M harmonic intervals in total, indexed by i):

$$\text{MAHIE} = \frac{1}{M} \sum_{i=1}^{M} |e_i^h|. \tag{2.9}$$

2.3.3 Pitch Variation

The pitch variation of a note is defined as the mean square pitch difference of the note trajectory from its median value. It indicates the extent of pitch variation over the duration of the note. The larger the pitch variation, the less stable the pitch. For a single note with N sampling points, where $p(i)$ represents the pitch at sampling point i and $\bar{p}$ is the median of $p(i)$ over the N points, the pitch variation V is calculated as follows:

$$v_i = \frac{1}{N} \sum_{i=1}^{N} |p(i) - \bar{p}|^2, \tag{2.10}$$

The *mean pitch variation* (MPV) is the mean value of pitch variation over multiple notes.

$$\text{MPV} = \frac{1}{M} \sum_{i=1}^{M} v_i \tag{2.11}$$

2.4 Pitch Drift

A Capella ensembles frequently observe a change in tuning over the duration of a piece, even when they are unable to detect any local changes. This phenomenon, called *intonation drift* or *pitch drift* (Seaton et al., 2013), usually exhibits as a lowering of pitch, or downward drift (Alldahl, 2006).

Different from pitch error and interval error, drift is not particular to specific notes, but

corresponds to a cumulative effect over a whole piece of music. Several studies have investigated pitch drift in unaccompanied singing (e.g. Howard, 2003; Terasawa, 2004; Kalin, 2005; Devaney and Ellis, 2008b; Mauch et al., 2014). Howard (2007b) tested the hypothesis that the use of just intonation, where the fundamental frequencies of pairs of simultaneous or consecutive notes are related by ratios of small whole numbers (Lindley, 2001), causes pitch drift. The hypothesis in such work is that the pitch adjustments required to intone pure intervals accumulate over time resulting in a shifting tonal reference (Mullen, 2000). Howard's study confirmed that singers make use of non-equal-tempered intonation to govern their tuning, and showed that it is possible to predict the direction of pitch drift in controlled harmonic progressions.

Mauch et al. (2014) defined the pitch drift as the mean pitch difference between corresponding tones in each repetition of the same piece (as shown in Equation 2.12).

$$D_{jk} = \frac{1}{M} \sum_{k=1}^{M} \left(e_{i+M \times (j-k)} - e_i \right) \tag{2.12}$$

where D_{jk} is the drift between j^{th} and k^{th} repetition, M is the number of notes in each repetition, and i is the index of the note. The term D_{jk} contains the magnitude and the direction of drift.

2.5 Relevant Tools

In this project, the software Tony (Mauch et al., 2015) was used as our annotation tool and MATLAB (MathWorks Core, R2009a) for data analysis. Tony was designed at QMUL's Centre for Digital Music in 2013 and is based on the pYIN algorithm (Mauch and Dixon, 2014). It takes monophonic recordings as input and extracts pitch information, segmenting the pitch track into notes. It allows users to annotate and play pitches alongside the recording. Users can merge, split, form, shift and delete the notes manually if they find errors. Then the data can be explored as .ton format or as .csv format for further processing. Tony default output is in frequency, thus the data was all transferred into MIDI notation according to the Equation 2.1.

Except for Chapter 3 which uses the YIN algorithm (de Cheveigné and Kawahara, 2002), all the rest of the experiments use pYIN to extract the fundamental frequency of any obtained recordings. The YIN algorithm features a very low error rate and few tuning parameters (de Cheveigné and Kawahara, 2002); while pYIN (Mauch and Dixon, 2014), is a probabilistic extension of YIN which provides robustness against errors due to sub-
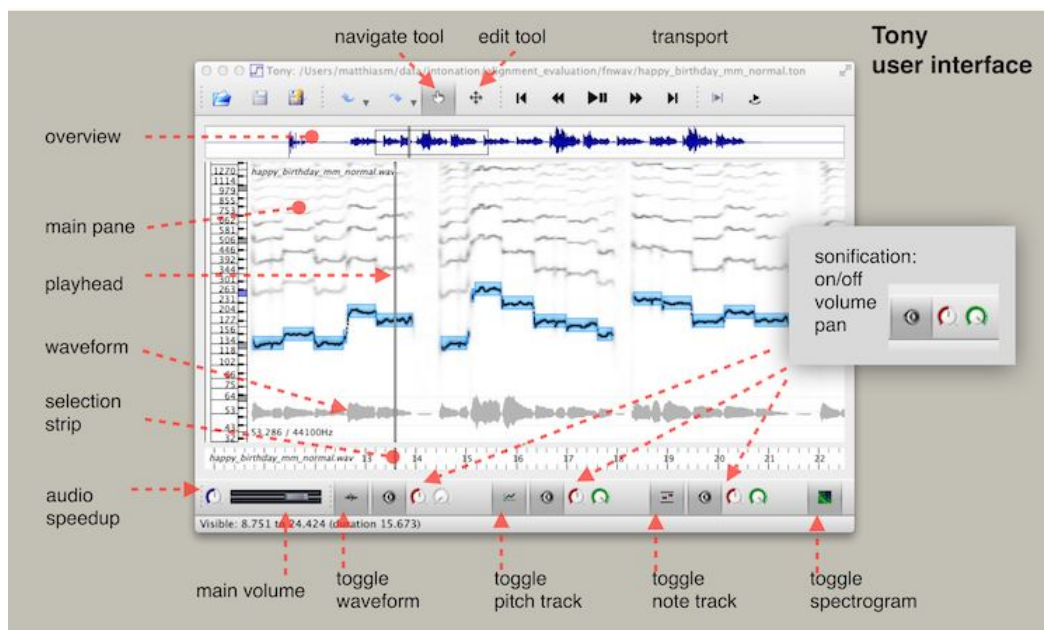
Figure 2.5: Interface and instruction of Tony.

optimal threshold settings.

2.6 Preliminary Project

This experiment was published in the first year of my PhD, although the data was collected for my master's thesis, the analysis and results are completely new compared with the masters thesis. A new dataset was created for singing analysis and modelling, and an exploratory analysis presented of pitch accuracy and pitch trajectories.

The aim of this preliminary project was to observe the vocal notes, to model the sung notes and investigate the correlation between musical background, melodic interval error and pitch accuracy. In this project, shortened versions of three pieces from *The Sound of Music* were sung three times by 39 participants without accompaniment, resulting in a dataset of 21762 notes in 117 recordings. Pitch estimates were obtained by the *Tony* software's automatic transcription and manual correction tools, and pitch accuracy is reported in terms of pitch error and interval error.

2.6.1 Musical material

Three songs were chose from the musical "The Sound of Music" as our material: "Edelweiss", "Do-Re-Mi" (shown in Figure 2.6) and "My Favourite Things." Despite originating from one work, the pieces were selected as being diverse in terms of tonal material and tempo (Table 2.1), well-known to many singers, and yet sufficiently challenging for ama-

Figure 2.6: Score of piece Do-Re-Mi, with some intervals marked (see Section 2.6.5)

Table 2.1: Summary details of the three songs used in this study.

Title	Tempo (BPM)	Key	Notes
Edelweiss	80	B♭	54
Do-Re-Mi	120	C	59
My Favourite Things	132	Em	73

teur singers. The pieces were shortened so as to contain a single verse without repeats, which the participants were asked to sing to the syllable /ta/ (shown in Appendix B). In order to observe long-term pitch trends, each song was sung three times consecutively. Each trial lasted a little more than 5 minutes.

2.6.2 Participants

39 participants were recruited (12 male, 27 female), most of whom are members of the university's music society or our music-technology focused research group. Some participants took part in the experiments remotely. The age of the participants ranged from 20 to 27 years (mean 23.3, median 23 years). All participants were asked to self-assess their musical background with questions loosely based on the Goldsmiths Musical Sophistication Index (Müllensiefen et al., 2011, 2014). Table 2.2 shows the results, suggesting a range of skill levels, with a strong bias towards amateur singers.

Table 2.2: Self-reported musical experience

Musical Background		Instrumental Training	
None	5	None	5
Amateur	27	1–2 years	15
Semi-professional	5	3–4 years	7
Professional	2	5+ years	12
Singing Skill		Singing Practice	
Poor	2	None	4
Low	25	Occasionally	22
Medium	9	Often	12
High	3	Frequently	1

2.6.3 Recording procedure and annotation

Participants were asked to sing each piece three times on the syllable /ta/. They were given the starting note but no subsequent accompaniment, except unpitched metronome clicks.

The software *Tony* was used to annotate the notes in the audio files (Mauch et al., 2015): pitch track and notes were extracted using the pYIN algorithm (Mauch and Dixon, 2014) and then manually checked and, if necessary, corrected. Approximately 28 corrections per recording were necessary.

2.6.4 Tonal reference curves and pitch error

In unaccompanied singing, pitch error is ill-defined, since singers use intonation with respect to their internal reference, which is very hard to track directly. If it is assumed that this internal reference doesn't change, pitch error can be estimated via the mean error with respect to a nominal (or given) reference pitch. However, it is well-known that unaccompanied singers (and choirs) do not maintain a fixed internal reference. For short pieces, this has been addressed by estimating the singer's reference frequency using linear regression (Mauch et al., 2014), but as there is no good reason to assume that drift is linear, a sliding window approach was proposed in order to provide a local estimate of tuning reference.

The first step is to take the annotated musical pitches p_i of a recording and remove the nominal pitch p_i^s given by the score, $t_i^* = p_i - p_i^s$, then apply further adjustment by subtracting the mean: $t_i = t_i^* - \bar{t}^*$. The reason for subtracting the mean is because

some of the participants did not start with the reference pitch. For example, one female participant sang one octave higher than the score pitch. Without subtracting the mean (about 12 semitones), all the pitch errors are about 12 semitones. The resulting raw tonal reference estimates t_i are then used as a basis for the tonal reference curves and pitch error calculations.

The second step is to find a smooth trajectory based on these raw tonal reference estimates. For each note, the weighted average of t_i was calculated by the mean of t_i in a context window around the note, obtaining the reference pitch c_i, from which the pitch error can be calculated:

$$c_i = \sum_{k=-n}^{n} w_k t_{i+k}, \tag{2.13}$$

where $\sum_{k=-n}^{n} w_k = 1$. Any window function $W = \{w_k\}$ can be used in Equation 2.13. Experiments were performed with symmetric windows with two different window shapes (rectangular and triangular) and seven window sizes (3, 5, 7, 9, 11, 15 and 25 notes) to arrive at smooth tonal reference curves. The rectangular window $W^{R,N} = \{w_k^{R,N}\}$ centred at the ith note is used to calculate the mean of its N-note neighbourhood, giving the same weight to all notes in the neighbourhood, but excluding the ith note itself:

$$w_k^{R,N} = \begin{cases} \frac{1}{N-1}, & 1 \leqslant |k| \leqslant \frac{N-1}{2} \\ 0, & \text{otherwise.} \end{cases} \tag{2.14}$$

The triangular window $W^{T,N} = \{w_k^{T,N}\}$ gives more weight to notes near the ith note (while still excluding the ith note itself). For example, if the window size is 5, then the weights are proportional to 1, 2, 0, 2, 1. More generally:

$$w_k^{T,N} = \begin{cases} \frac{2N+2-4|k|}{N^2-1}, & 1 \leqslant |k| \leqslant \frac{N-1}{2} \\ 0, & \text{otherwise.} \end{cases} \tag{2.15}$$

The smoothed tonal reference curve c_i is the basis for calculating the pitch error:

$$e_i^p = t_i - c_i, \tag{2.16}$$

The tonal reference curves c_i can also be used to calculate a new measure of the extent of fluctuation of a singer's reference pitch. This measurement was called the tonal reference
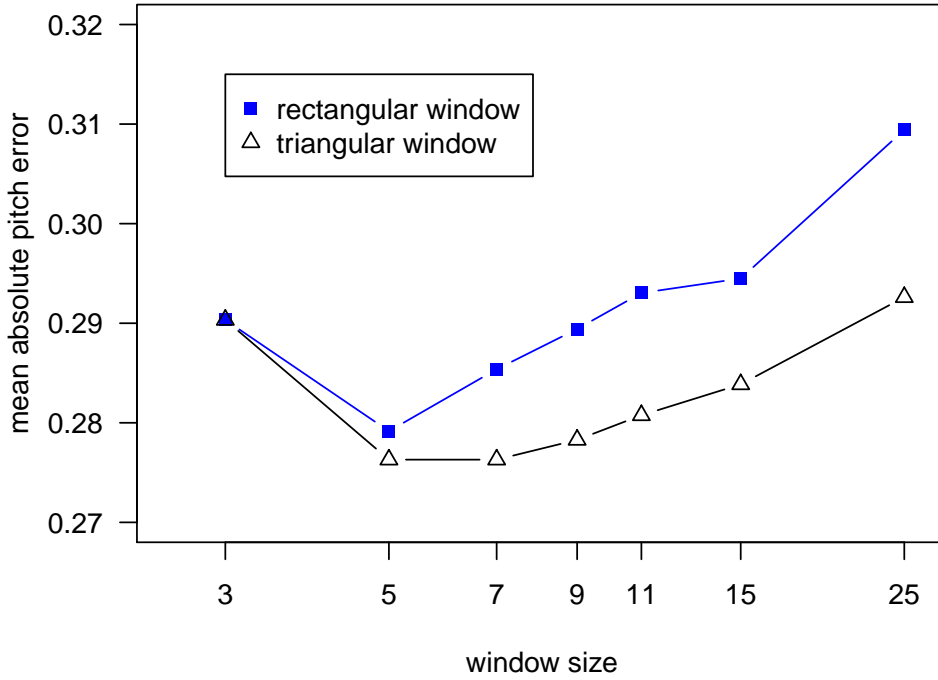
Figure 2.7: Pitch error (MAPE) for different sliding windows.

deviation (TRD), calculated as the standard deviation:

$$\text{TRD} = \sqrt{\frac{1}{M-1} \sum_{i=1}^{M} (c_i - \bar{c}_M)^2}. \tag{2.17}$$

2.6.5 Results

Firstly, multiple choices of window size and shape are compared for the calculation of the smoothed tonal reference curves c_i (Equation 2.13 and 2.17), which provide the local tonal reference estimate used for calculating mean absolute pitch error (MAPE). It was assumed that the window that gives rise to the lowest MAPE models the data best. Figure 2.7 shows that for both rectangular and triangular window shapes an intermediate window size N of 5 notes minimises MAPE, with the triangular window working best (MAPE = 0.276 semitones, computed over all singers and pieces). Hence, this window is used for computations relating to pitch error, including tonal reference curves, and for understanding how pitch error is linked to note duration and singers' self-reported skill and experience, which is not used in other Chapters.

Smoothed tonal reference curves

The smoothed curves exhibit some unexpected behaviour. Figure 2.8 shows three examples of different participants and pieces. Several patterns emerge. Figure 2.8a shows a performance in which pitch error is kept within half a semitone and tonal reference is almost completely stable. This is reflected in very low values of MAPE (0.171) and TRD (0.070), respectively.

However, most singers' tonal reference curves fluctuate. For example, Figure 2.8b illustrates a tendency of some singers to smoothly vary their pitch reference in direct response to the piece. The trajectory shows a periodic structure synchronised with the three repetitions of the piece. The fluctuation measure TRD is much higher as a result (0.624). This is a common pattern which was observed.

The third example (Figure 2.8c) illustrates that strong fluctuations are not necessarily periodic. Here, TRD (0.635) is nearly identical, but originates from a mostly consistent downward trajectory. The singer makes significant errors in the middle of each run of the piece, most likely due to the difficult interval of a downward tritone occurring twice (notes 42 and 50; more discussion below). Comparing Figures 2.8b and 2.8c also shows that MAPE and TRD are not necessarily related. Despite large fluctuations (TRD) in both, pitch error (MAPE) is much smaller in Figure 2.8c (0.297).

Turning from the trajectories to pitch error measurements, the three pieces show distinct patterns (Figure 2.9). The first piece, Edelweiss, appears to be the easiest to sing, with relatively low median pitch errors. In Do-Re-Mi, the third quarter of the piece appears much more difficult than the rest. This is most likely due to faster runs and the presence of accidentals, taking the singer out of the home tonality. Finally, My Favourite Things exhibits a very distinct pattern, with relatively low pitch errors throughout, except for one particular note (number 50), which is reached via a downward tritone, a difficult interval to sing. The same tritone (A-D♯) occurs at note 42, where the error is smaller and notably in the opposite direction (this D♯ is flat, while note 50 is over a semitone sharp on average). It appears that singers are drawn towards the more consonant (and more common) perfect fifth and fourth intervals, respectively.

Duration, interval and proficiency factors

The observations on pitch error patterns suggest that note duration and the tritone interval may have significant impact on pitch error. In order to investigate their impact, a linear model was adopted, taking into account furthermore the size of the intervals sung and
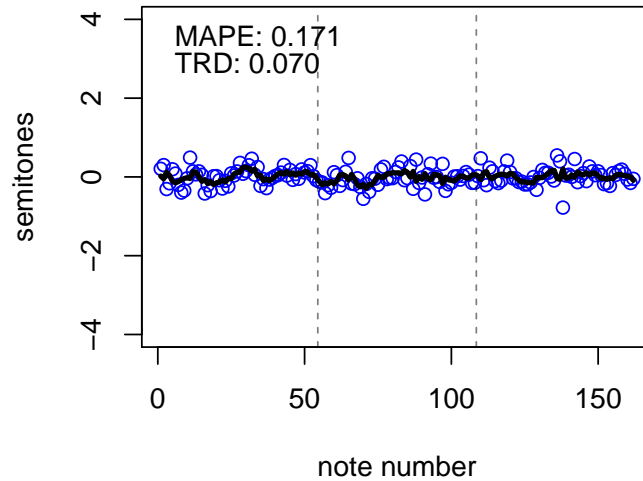
(a) Edelweiss, singer 11
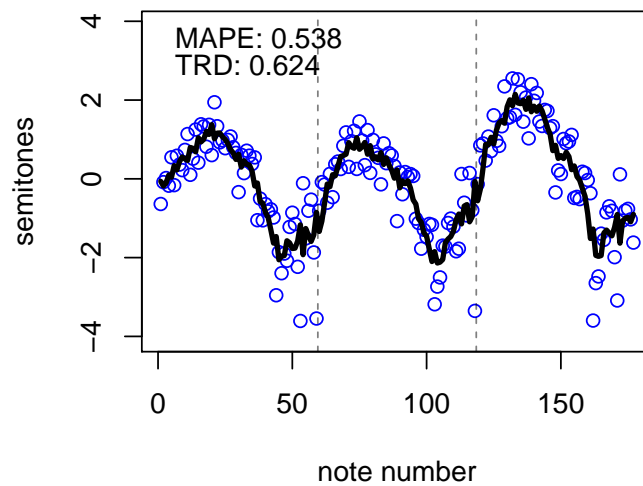
(b) Do-Re-Mi, singer 39
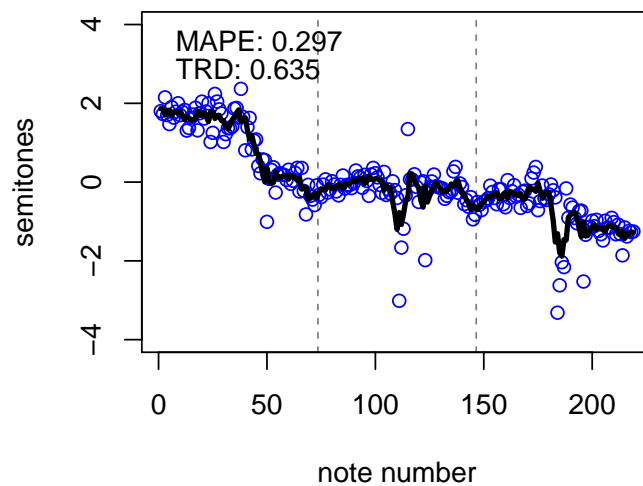
(c) My Favourite Things, singer 31

Figure 2.8: Examples of tonal reference trajectories. Dashed vertical lines delineate the three repetitions of the piece where the blue circles are the observed pitch errors, and the black line is the TRD.

(a) Edelweiss

(b) Do-Re-Mi

(c) My Favourite Things
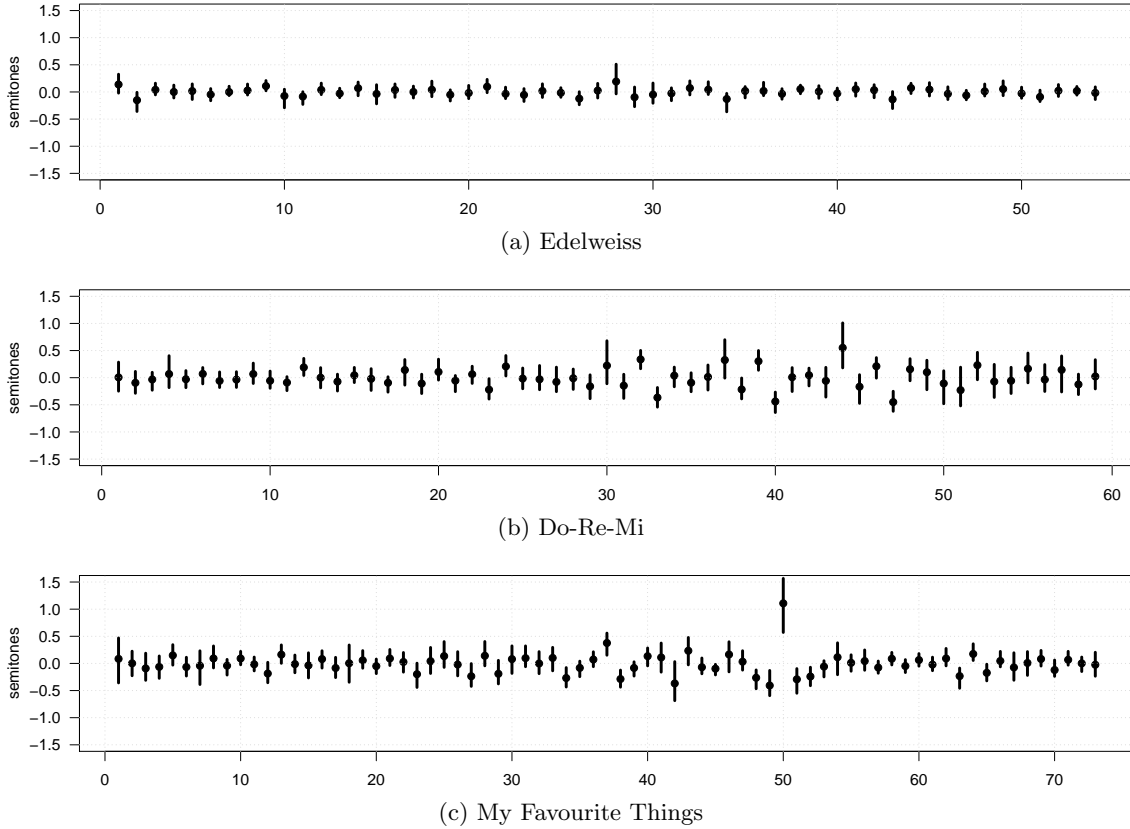
Figure 2.9: Pitch errors by note index for each of the three pieces. The plots show the median values with bars extending to the first and third quartiles.

singer bias via considering the singers' self assessment.

Table 2.3a lists all dependent variables, estimates of their effects and indicators of significance. The following discussion describes how these variables influence, reduce or add to error, but note that the model gives no indication of true causation, only of correlation. The first question is whether note duration influences pitch error. The intuition is that longer notes, and notes with a longer preparation time (previous inter-onset interval, IOI), should be sung more correctly. This is indeed the case. The results observe a reduction of pitch error of 0.073 semitones per added second of duration. The IOI between the previous and current note also reduces pitch error, but by a smaller factor (0.021 semitones per second). Conversely, absolute nominal interval size adds to absolute pitch error, by about 0.016 semitones per interval-semitone, as does the absolute size of the next interval (0.010 semitones). The intuition about the tritone interval is confirmed here, as the presence of any tritone (whether upward or downward) adds 0.370 semitones—on average—to the absolute pitch error. The last covariate, questionnaire score, is the sum of the points obtained from the four self-assessment questions, with values ranging between 5 and 14. The result shows that there is correlation between the singers' self-assessment and their

	Estimate	Std. Err.	t	p
(intercept)	0.374	0.012	32.123	0.000
nominal duration	-0.073	0.004	-17.487	0.000
prev. nom. IOI	-0.021	0.004	-4.646	0.000
abs(nom. interval)	0.016	0.001	13.213	0.000
abs(next nom. interval)	0.010	0.001	8.471	0.000
tritone	0.370	0.019	19.056	0.000
quest. score	-0.011	0.001	-9.941	0.000

(a) MAPE

	Estimate	Std. Err.	t	p
(intercept)	0.481	0.015	33.124	0.000
nominal duration	-0.076	0.005	-14.570	0.000
prev. nom. IOI	-0.050	0.006	-8.984	0.000
abs(nom. interv.)	0.030	0.002	19.700	0.000
abs(next nom. interv.)	-0.006	0.002	-3.826	0.000
tritone	0.373	0.024	15.404	0.000
quest. score	-0.012	0.001	-8.665	0.000

(b) MAMIE

Table 2.3: Effects of multiple covariates on error for a linear model. t denotes the test statistic. The p value rounds to zero in all cases, indicating statistical significance.

absolute pitch error. For every additional point in the score their absolute pitch error is reduced by 0.012 semitones. The picture is very similar for the analysis of absolute interval error (Table 2.3b): the effect directions of the variables are the same, except for the size of the following interval, where the effect size is smallest.

2.6.6   Discussion

This experiment has investigated how note length relates to singing accuracy, finding that notes are sung more accurately as the singer has more time to prepare and sing them. Yet it is not entirely clear what this improvement is based upon. Chapter 4 has similar results where duration significantly influences the intonation accuracy. Do longer notes genuinely give singers more time to find the pitch, or is part of the effect due to measurement or statistical artefacts? Evidence could be found by examining pitch at the sub-note level, taking vibrato and note transitions into account. Conversely, studying the effect of melodic context on the underlying pitch track could shed light on the physical process of singing, and could be used for improved physical modelling of singing.

Overall, the absolute pitch error of singers (mean: 28 cents; median: 18; SD: 36) and the absolute interval error (mean: 34 cents; median: 22; SD: 46) are slightly higher than those reported elsewhere (Mauch et al., 2014), but this may reflect the greater difficulty of our musical material in comparison to "Happy Birthday". The analysis also did not exclude singers with large pitch errors, although the least accurate singers had MAPE and MAMIE values of more than half a semitone, i.e. they were on average closer to an erroneous note than to the correct one. That the values of MAMIE and MAPE are similar is to be expected, as interval error is the limiting case of pitch error, using a minimal window containing only the current and previous note.

A symmetric window was used in this work, but this could easily be replaced with a causal (one-sided) window (Mauch et al., 2014), which would also be more plausible psychologically, as the singer's internal pitch reference in the model (Equation 2.13 to 2.15) is based equally on past sung notes and future not-yet-sung notes. However, for post hoc analysis, the fuller context might reveal more about the singer's internal state (which must influence the future tones) than the more restricted causal model.

Figure 2.9 shows how the three pieces differ in terms of pitch accuracy. It is interesting to see that accidentals (which result in a departure from the established key), and the tritone as a particular example, seem to have a strong adverse impact on accuracy. To compile more detailed statistical analyses like the ones in Table 2.3 one could conduct

singing experiments on a wider range of intervals, isolated from the musical context of a song.

Finally, note that some singers took prolonged breaks between runs in a three-run rendition of a song. The recording was stopped, but no new reference note was played, so the singers resumed with the memory of what they last sung. As part of the reproducible code package, information is provided on which recordings were interrupted and at which point. The results show that the regression coefficients (Table 2.3) did not substantially change as a result of these interruptions.

2.6.7 Conclusions

This preliminary project presented a new dataset for singing analysis, investigating the effects of singer and piece factors on the intonation of unaccompanied solo singers. Pitch accuracy was measured in terms of pitch error and interval error. By introducing a new model of tonal reference computed using the local neighbourhood of a note, it was found that a window of two notes each side of the centre note provides the best fit to the data in terms of minimising the pitch error. The temporal evolution of tonal reference during a piece revealed patterns of tonal drift in some singers, others appeared random, yet others showed periodic structure linked to the score. As a complement to errors in individual notes or intervals, a measure was introduced for the magnitude of drift, tonal reference deviation (TRD), and its behaviour illustrated using several examples.

Two types of factors influencing pitch error were investigated, those related to the singers and those related to the material being sung. In terms of singer factors, results show that pitch accuracy correlates with self-reported singing skill level, musical training, and frequency of practice. Larger intervals in the score led to larger errors, but only accounted for 2–3 cents per semitone of the mean absolute errors. On the other hand, the tritone interval accounted for 37 cents of error when it occurred, and in one case led to a large systematic error across many of the singers. Results also indicate that note duration has an effect on pitch accuracy, as singers make use of aural feedback to regulate their pitch, which results in less stable pitch at the beginnings of notes. A small but significant effect of duration was found for both the current note, and the nominal time taken from the onset of the previous note; longer duration led to greater accuracy.

2.6.8   Data Availability

All audio recordings and corresponding trajectory plots for this experiment can be obtained from `http://dx.doi.org/10.6084/m9.figshare.1482221`.

The code and the data needed to reproduce the results (note annotations, questionnaire results, interruption details) are provided in an open repository at `https://code.soundsoftware.ac.uk/projects/dai2015analysis-resources`.

Chapter 3

Imitation Study

The motivation of this chapter is to investigate the correlation between pitch parameters and vocal accuracy. A total of 43 selected singers were asked to vocally imitate a set of stimuli, where these stimuli were designed and grouped for the purpose of parameter control and classification. Five types of stimulus were used: a constant pitch (*stable*), a constant pitch preceded by a pitch glide (*head*), a constant pitch followed by a pitch glide (*tail*), a pitch *ramp* and a pitch with *vibrato*; with parameters for main pitch, transient length and pitch difference. Besides stimuli type, two conditions were tested: singing simultaneously with the stimulus, and singing alternately, between repetitions of the stimulus. After automatic pitch-tracking and manual checking of the data, intonation accuracy and variance were calculated. The pitch trajectories of observed data were modelled by the shapes of given stimuli and the parameters were optimised. The resulting parameters matched the stimuli more closely than the raw data did. Pitch error was modelled with a linear mixed-effects model, and factors were tested for significant effects using one-way analysis of variance. The results indicate: (1) Significant factors include stimulus type, main pitch, repetition, condition and musical training background, while order of stimuli, sex and age do not have any significant effect. (2) The *ramp*, *vibrato* and *tail* stimuli have significantly greater absolute pitch errors than the *stable* and *head* stimuli. (3) Pitch error shows a small but significant linear trend with pitch difference. (4) Notes with shorter transient duration are more accurate. These results offer a better understanding of pitch accuracy with unstable accompaniment which supports the design of further experiments.

3.1 Research Questions and Hypotheses

Before the investigation of the complex interaction between singing ensembles, a first experiment was designed to explore how individual singers respond to pitch variation. Previous research has highlighted significant differences when singers sing in unison or individually, but any specific correlations between stimuli and pitch still need further investigation. To study the response of singers to time-varying pitch trajectories, a controlled experiment was presented by using synthetic stimuli, in order to test the following hypotheses:

1. The stimulus type will have a significant effect on intonation accuracy.

2. A greater duration or extent of deviation from the main pitch will increase intonation error.

3. The direction of any deviation in the stimulus from the main pitch determines the direction of any error in the response.

4. Singing simultaneously with the stimulus will result in a lower error than alternating the response with the stimulus.

The first hypothesis is based on in previous studies (Chapter 2) which suggest that singers are distracted by simultaneous sounds when they are singing. To investigate how simultaneous sounds influence the pitch, stimuli are chosen with different types and parameters, including the variation in transient parts, the extent of pitch deviation from the main pitch, the duration of transient parts, the direction of the transient parts, and singing conditions. The experiment supposes that the type of the stimuli and numerical differences in these parameters affect intonation accuracy significantly. Thus as a corollary, it could be expected that more accurate intonation when stimuli are close to constant pitch. This leads to the hypotheses: greater pitch variation inside a stimulus increases the pitch error (hypothesis 2); the direction of the stimulus has an effect on the direction of pitch error (hypothesis 3).

When participants sing simultaneously, they can adjust the pitch immediately with the stimulus, while they have to rely on their tonal memory to adjust to the correct pitch when singing alternately. It might be harder to keep in tune without an alignment. Thus the last hypothesis assumes that intonation is more accurate in the simultaneous condition than in the alternating condition (hypothesis 4).

The confirmation or refutation of the hypotheses shows the existence of the interaction between acoustic input and singing output, and helps to inform the subsequent experiments.

3.2 Experiment Design

According to the hypotheses and the observations of the previous work (Section 2.6), five stimulus types were selected representing the basic shapes of vocal notes: a constant pitch, a note with a transient part at the beginning, a transient part at the end, a vibrato note, and a note with pitch gradually changing. For each type of stimulus, various parameters were set according to the previous observations. The combinations of types and parameters led to 75 different individual stimuli (Table 3.1). All the stimuli were repeated in each trial while all the trials were tested in both listening conditions (simultaneously or alternatively).

In each trial, the participant imitated the stimulus three times (see Figure 3.1). Each stimulus was one second in duration. In the *simultaneous* condition, the stimulus was repeated six times, with one second of silence between the repetitions, and the participants sang simultaneously with the 2nd, 4th and 6th instances of the stimulus. The *sequenced* condition was similar in that the responses occurred at the same times as in the simultaneous case, but the stimulus was not played at these times. There was a three second pause after each trial. The trials of a given condition were grouped together, and participants were given visual prompts so that they knew when to respond. Each of the 75 trials within a condition used a different stimulus, taken from one of the five stimulus types described in Section 3.2.1, and presented in a random order. The two conditions were also presented in a random order.

3.2.1 Stimuli

Unlike previous imitation experiments which have used fixed-pitch stimuli, the experimental stimuli were synthesised from time-varying pitch trajectories in order to provide controlled conditions for testing the effect of specific deviations from constant pitch. Five stimulus types were chosen, representing a simplified model of the components of sung tones (constant pitch, initial and final glides, vibrato and pitch ramps). The pitch trajectories of the stimuli were generated from the models described below and synthesised by a custom-made MATLAB program, using a solo male voice on the vowel /a:/. Due to the limitation of the synthesis package, all the stimuli use the male voice.
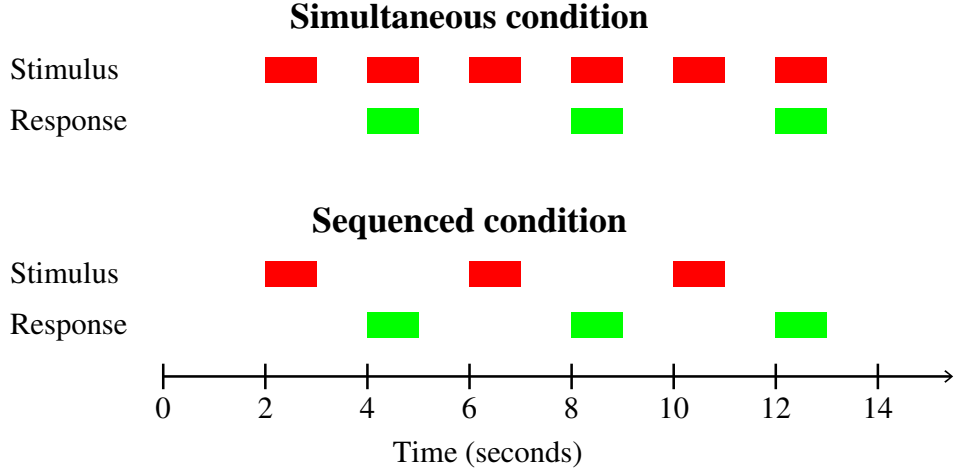
Figure 3.1: Experimental design showing timing of stimuli and responses for the two conditions.

Table 3.1: Parameter settings for each stimulus type. The columns contain stimulus types, main pitch p_m, duration d and pitch difference p_D in semitone and the count of stimuli. The octave for the pitch parameter was dependent on sex (3 for male, 4 for female).

Type	p_m	d	p_D	Count
stable	$\{C, F, B\flat\}$	$\{0.0\}$	$\{0.0\}$	3
head	$\{C, F, B\flat\}$	$\{0.1, 0.2\}$	$\{\pm 1, \pm 2\}$	24
tail	$\{C, F, B\flat\}$	$\{0.1, 0.2\}$	$\{\pm 1, \pm 2\}$	24
ramp	$\{C, F, B\flat\}$	$\{1.0\}$	$\{\pm 1, \pm 2\}$	12
vibrato	$\{C, F, B\flat\}$	$\{\pm 0.32\}$	$\{0.25, 0.5\}$	12

The five different stimulus types considered in this work are: constant pitch (*stable*), a constant pitch preceded by an initial quadratic pitch glide (*head*), a constant pitch followed by a final quadratic pitch glide (*tail*), a linear pitch ramp (*ramp*), and a pitch with sinusoidal vibrato (*vibrato*). The stimuli are parameterised by the following variables: p_m, the main or central pitch; d, the duration of the transient part of the stimulus; and p_D, the extent of pitch deviation from p_m. For *vibrato* stimuli, d represents the period of vibrato. Values for each of the parameters are given in Table 3.1 and the text below.

By assuming an equal tempered scale with reference pitch A4 tuned to 440 Hz, pitch p in semitones and fundamental frequency F_0 can be related as given in Equation 2.1 (Mauch et al., 2014).

For the *stable* stimulus, the pitch trajectory $p(t)$ is defined as follows:

$$p(t) = p_m, \qquad 0 \leqslant t \leqslant 1. \tag{3.1}$$

The *head* stimulus is represented piecewise by a quadratic formula and a constant:

$$p(t) = \begin{cases} at^2 + bt + c, & 0 \leqslant t \leqslant d \\ p_m, & d < t \leqslant 1. \end{cases} \tag{3.2}$$

The parameters a, b and c are selected to make the curve pass through the point $(0, p_m + p_D)$ and have its vertex at (d, p_m), where $a = p_D/d^2$, $b = -2 \times p_D/d$ and $c = p_m + p_D$. The *tail* stimulus is similar, with $p(t) = p_m$ for $t < 1 - d$, and the transient section being defined for $1 - d \leqslant t \leqslant 1$. In this case the parameters a, b and c are chosen so that the curve has vertex $(1-d, p_m)$ and passes through the point $(1, p_m + p_D)$, where $a = p_D/d^2$, $b = 2 \times p_D(d-1)/d^2$ and $c = p_m + p_D(d - 1/d)^2$.

The *ramp* stimuli are defined by:

$$p(t) = p_m + p_D \times (t - 0.5), \qquad 0 \leqslant t \leqslant 1. \tag{3.3}$$

Finally, the equation of *vibrato* stimuli is:

$$p(t) = p_m + p_D \times \sin\left(\frac{2\pi t}{d}\right), \qquad 0 \leqslant t \leqslant 1. \tag{3.4}$$

There is a substantial amount of data on the fundamental frequency of the voice in the speech of speakers who differ in age and sex (Traunmüller and Eriksson, 1995). Three pitch values were chose according to sex to fall within a comfortable range for most singers. The pitches C3 ($p_m = 48$), F3 ($p_m = 53$) and B♭3 ($p_m = 58$) were chosen for male singers and C4 ($p_m = 60$), F4 ($p_m = 65$) and B♭4 ($p_m = 70$) for female singers. For the *vibrato* stimuli, the vibrato rates were selected as 3 and 6 Hz according to a reported mean vibrato rate across singers of 6.1 Hz (Prame, 1994), and the extent or depth of vibrato to ±0.25 or ±0.5 semitones, in accordance with values reported by Sundberg (1994). Because intonation accuracy is affected by the duration of the note (Fyk, 1985; Dai et al., 2015), a fixed one-second duration was used for all stimuli in this experiment. Figure 3.2 shows the visualised stimulus shapes.

3.3 Implementation

For attracting sufficient participants and managing the received data, an on-line test system was adapted for remote data collection and a management system for data storage. Singers only needed a laptop with an earphone and microphone to participate in the
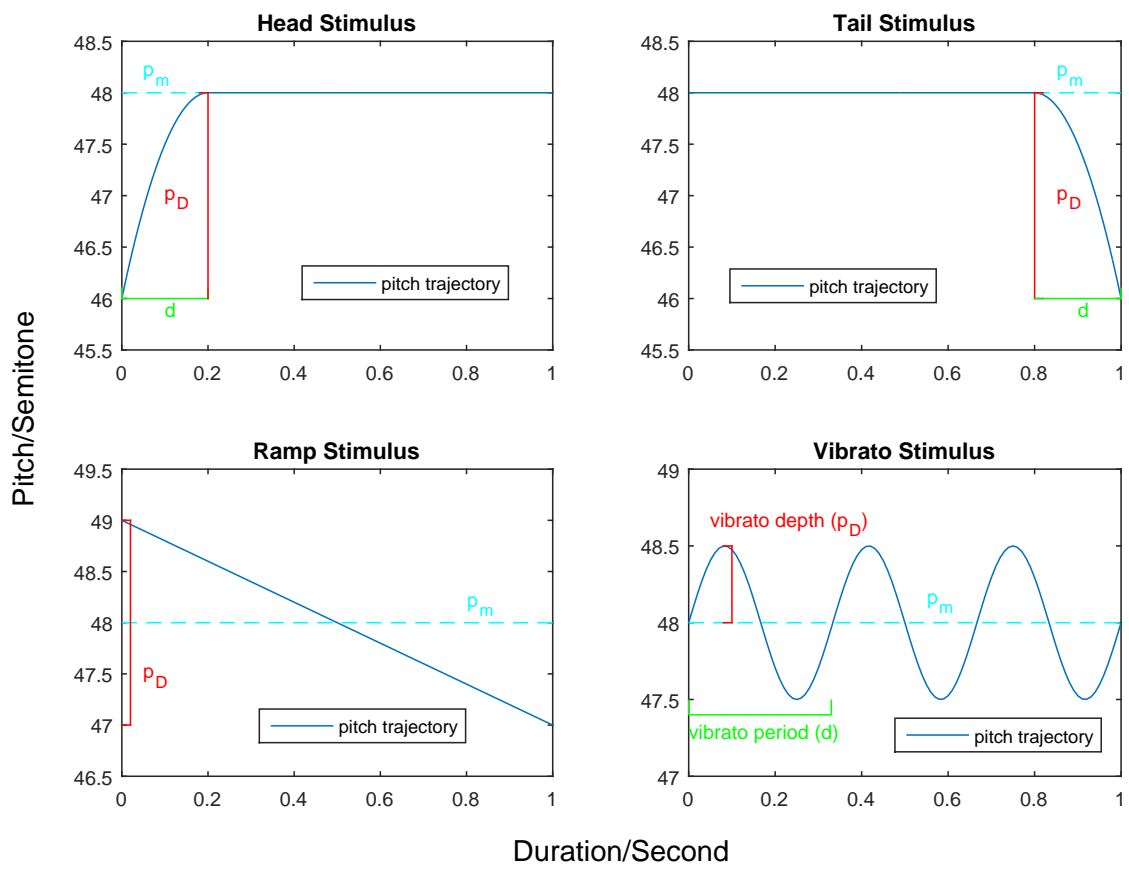
Figure 3.2: Stimulus types and parameters for the four non-constant types.

experiments, although most attended the research lab to take part. There is a tutorial on YouTube.com for guiding the participants on how to finish the experiment on-line (link can be found in Section 3.8).

3.3.1 Participants

A total of 43 participants (27 female, 16 male) took part in the experiment. 38 of them were recorded in the studio and 5 were distance participants from the USA, Germany, Greece and China (2 participants). The range of ages was from 19 to 34 years old (mean: 25.1; median: 25; SD: 2.7). Apart from 3 participants who did not complete the experiment, all the other singers recorded all the trials.

The researcher intentionally chose non-poor singers as the research target. "Poor-pitch singers" are defined as those who have a deficit in the use of pitch during singing (Welch, 1979; Pfordresher and Brown, 2007), and are thus unable to perform the experimental task. Participants whose pitch imitations had on average at least one semitone absolute error were categorised as poor-pitch singers. The data of poor-pitch singers is not included in this study (12 poor-pitch singers), apart from one singer who occasionally sang one octave higher than the target pitch.

Vocal training is an important factor for enhancing the singing voice and making the singer's voice different from that of an untrained person (Mendes et al., 2003). To allow us to test for the effects of training, participants completed a questionnaire containing 34 questions from the Goldsmiths Musical Sophistication Index (Müllensiefen et al., 2014) which can be grouped into 4 main factors for analysis: active engagement, perceptual abilities, musical training and singing ability (9, 9, 7 and 7 questions respectively as shown in Appendix C).

The study was conducted with the approval of the Queen Mary Ethics of Research Committee (approval number: QMREC1421).

3.3.2 Recording Procedure

A tutorial video was played before participation. In the video, participants were asked to repeat the stimulus precisely. They were not told the nature of the stimuli. Singers who said they could not imitate the time-varying pitch trajectory were told to sing a stable note of the same pitch.

The experimental task consisted of 2 conditions, each containing 75 trials, in which participants sang three one-second responses in a 16-second period. It took just over one hour

for participants to finish the experiment. 22 singers took the simultaneous condition first and 21 singers took the sequenced condition first. Although the synthetic stimulus simulated the vowel /a:/, participants occasionally chose other vowels that felt comfortable. The online system of data collection can be found in Appendix A which is a custom-made system specific for data collection of this experiment.

3.3.3 Extract Fundamental Frequency

The project in Section 2.6 used software Tony as the annotation tool (Mauch et al., 2015). However this experiment involved over 3000 audio files, so it was impossible to annotate them manually. Therefore, the researcher chose Sonic Annotator with pYIN Vamp plug-in as the F_0 extraction method (Mauch and Dixon, 2014). This extracted the timing and fundamental frequency information, and saved the data in a .csv file for further processing. The results were checked and it was found that 151 files are not extracted successfully among all the files. The unsuccessful files were detected automatically by their duration, pitch and segmentation results. Figure 3.3 (a) shows an example where F_0 was not extracted ideally. As shown in the figure, some pitch data is missing, and it is hard to model the note trajectory using the extracted data. Hence, Figure 3.3 (b) shows the result of improving the pitch detection method by adding band-pass filtering and power detection to the Yin algorithm (de Cheveigné and Kawahara, 2002).

The first step of improvement is the filtering of useful frequencies by a band-pass filter. The lowest note in this experiment has F_0 above 130 Hz and the highest note is about 466 Hz. The next step is analysis of the power information and filtering out the low power signal. The researcher randomly chose one or two files to set the power threshold while pitches with low power were omitted. This proved an effective method for blocking the noise. But for some notes, the improved Yin algorithm is not as accurate as the pYIN algorithm. Therefore, both methods were retained. For some particular notes, the researcher kept the results that displayed more accurate pitch extraction. The pYIN algorithm is more accurate than the traditional YIN algorithm, but it is difficult to modify, so these improvements are based on the YIN algorithm. The results of the improved method were greater continuity of pitch trajectories, as shown in Figure 3.3 (b). After improvement of the pitch detection, only five files were detected as unsuccessful. The 5 final unsuccessful files were annotated manually.
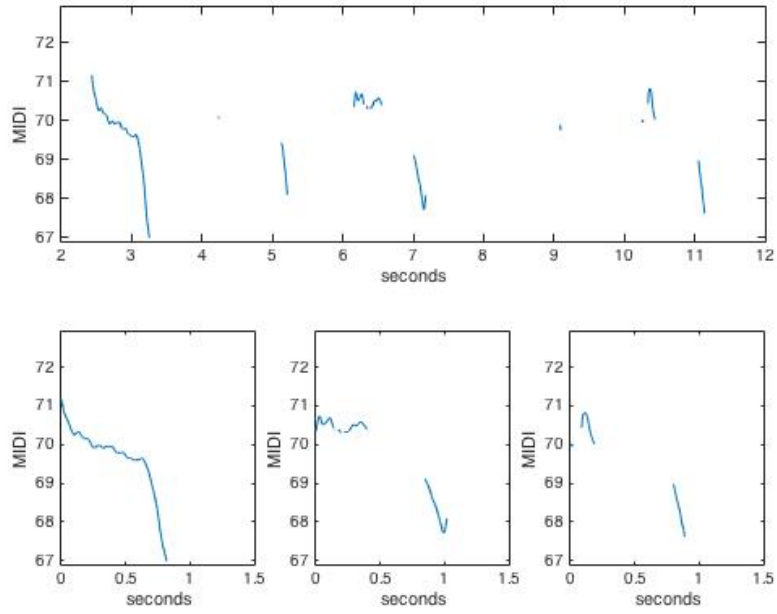
(a) F_0 extracted by pYIN
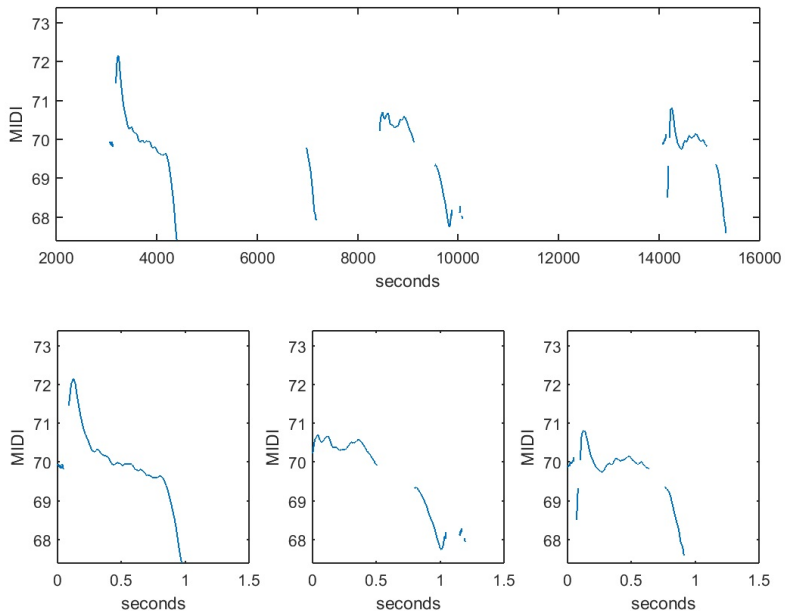
(b) F_0 extracted by YIN and power threshold

Figure 3.3: F_0 extracted by YIN and power threshold.

3.3.4 Segmentation

Every trial results in a 16 seconds recording containing 3 vocal notes. After extracting the F_0 , the recording is separated into three individual arrays.

The first step is to smooth the raw data by applying a smoothing window which calculates the mean of 5 adjacent sampling points, which can connect break-point inside the note. The smoothing window is usually shorter than 0.1 seconds depending on the sampling rate. The next step is application of a band pass filter, only retaining the data within ± 5 semitones of F_0, thereby getting rid of noise. Any remaining notes with a duration of less than 0.4 seconds were flagged and checked manually; most of these deficient notes were due to participants making no response. The onset and offset detection will find all the note durations from the power information. After sorting by duration, the longest three notes are kept. This is because in some recordings, participants sang more than 3 notes, or the pitch extraction recognised noise as pitch information.

3.3.5 Annotation

The main pitch p of the response is calculated by removing the first 10% and last 10% of the response duration according to the previous observation in the preliminary project, and computing the median of the remaining pitch track. The pitch error e^p is calculated as the difference between the main pitch of the stimulus p_m and that of the response p:

$$e^p = p - p_m \tag{3.5}$$

To avoid bias due to large errors, the researcher excluded any responses with $|e^p| > 2$ (4% of responses). Such errors arose for example when participants sang the pitch of the previous stimulus or one octave higher than the current stimulus. The resulting database contains 18572 notes, from which the statistics below are calculated.

The mean pitch error (MPE) over a number of trials measures the tendency to sing sharp (MPE > 0) or flat (MPE < 0) relative to the stimulus. The mean absolute pitch error (MAPE) measures the spread of a set of responses. These can be viewed respectively as inverse measures of accuracy and precision (cf. Pfordresher et al. (2010)).

To analyse differences between the stimulus and response as a time series, pitch error $e_f^p(t)$ is calculated frame-wise: $e_f^p(t) = p_r(t) - p_s(t)$, for stimulus $p_s(t)$ and response $p_r(t)$, where the subscript f distinguishes frame-wise results and t is time. For frame period T and frame index i, $0 \leqslant i < M$, the researcher calculates summary statistics:

$$\text{MAPE}_f = \frac{1}{M} \sum_{i=0}^{M-1} |e_f^p(iT)| \qquad (3.6)$$

and MPE_f is calculated similarly. MAPE_f is applied to investigate the mean absolute pitch error within notes. Equation 3.6 assumes that the two sequences $p_r(t)$ and $p_s(t)$ are time-aligned. Although cross-correlation could be used to find a fixed offset between the sequences, or dynamic time warping could align corresponding features if the sequences proceed at different or time-varying rates, in this case the researcher considered singing with the correct timing to be part of the imitation task, and the stimulus was aligned to the beginning of the detected response.

3.4 Results

Based on the previous research methods, the stimulus was modelled and extracted by several parameters such as fundamental frequency, frequency difference, vibrato depth and variation range.

3.4.1 Music Background

Vocal training has been promoted as a major factor for enhancing the singing voice. It is also a factor in developing those parameters that make a singer's voice different from that of a non-trained singer (Mendes et al., 2003). The participants performed a self-assessment of their musical background with questions from the Goldsmiths Musical Sophistication Index (Müllensiefen et al., 2011, 2014). The factor of music background was discussed in Section 2.2.5 and all the questions are in Appendix C. When compared with the MAPE, a t-test shows that all the factors are independent and have a different distribution. Figure 3.4 shows the results of the questionnaire.

Most of the participants are non-trained singers, but most of them have a few years of music training.

3.4.2 Influence of stimulus type on absolute pitch error

Pitch error (MPE: 0.0123; SD: 0.3374), absolute pitch error (MAPE: 0.2441; SD: 0.2332) and frame-wise absolute pitch error (MAPE_f: 0.3968; SD: 0.2238) were reported between all the stimuli and responses. 71.1% of responses have an absolute error less than 0.3 semitones. 51.3% of responses are higher (sharper) than the stimulus ($e^p > 0$). All the singers' information, questionnaire responses, stimulus parameters and calculated errors
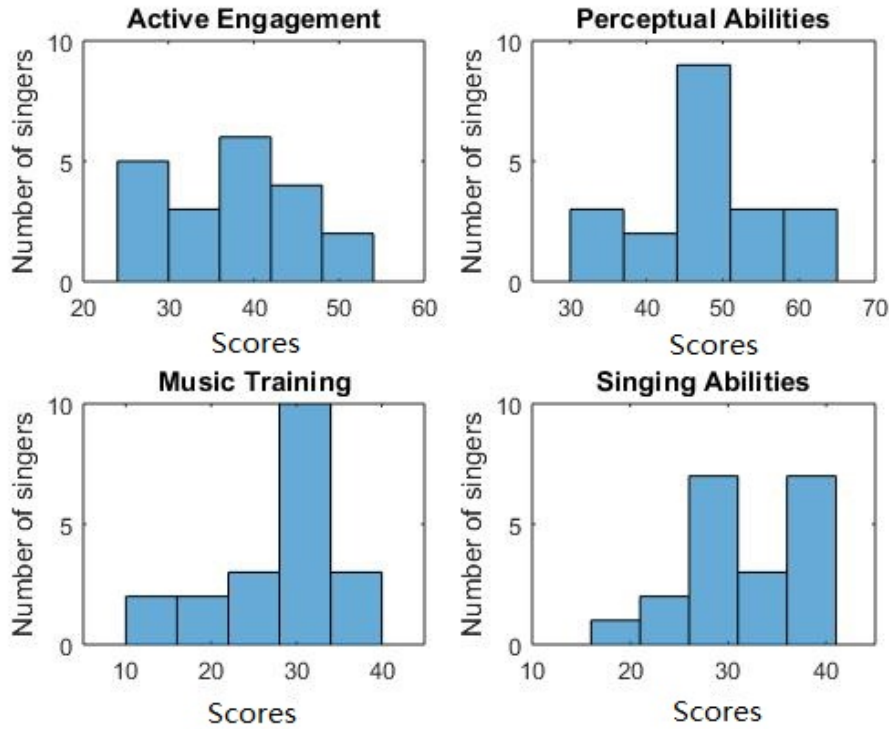
Figure 3.4: Histograms of questionnaire score for each category.

were arranged in a single table for further processing. The factors influencing absolute pitch error are analysed in this and the next subsection, and pitch error is considered in subsection 3.4.4, and the modelling of responses in Section 3.5.

One-way independent samples analysis of variance (one-way ANOVA) was performed with the fixed factor stimulus type (five levels: *stable*, *head*, *tail*, *ramp* and *vibrato*) and the random factor participant. There was a significant effect of stimulus type ($F(4, 18567) = 72.3$, $p < .001$), different stimulus types show a statistically significant effect on pitch accuracy.

Post hoc comparisons using the Tukey HSD test indicated that the absolute e^p for *tail*, *ramp* and *vibrato* stimuli were significantly different from those of the *stable* stimuli, while the *head* stimuli showed no significant difference from *stable* stimuli (see Table 3.2). Thus *tail*, *ramp* and *vibrato* stimuli do have an effect on pitch precision. Table 3.2 also shows the 95% confidence intervals for each stimulus type. Effect sizes were calculated by a linear mixed-effects model comparing with *stable* stimulus results.

The combined score is the sum of the other four factors from the questionnaires. For the effect of musical background, an ANOVA found that all background factors are significant for pitch accuracy (see Table 3.3). The task involved both perception and production, so it is to be expected that both of these factors (perceptual and singing abilities) would

60

Stimulus	MAPE	Confidence interval	Effect size
stable	0.1977	[0.1812, 0.2141]	–
head	0.1996	[0.1938, 0.2054]	0.2 cents
tail	0.2383	[0.2325, 0.2441]*	4.1 cents
ramp	0.3489	[0.3407, 0.3571]***	15.1 cents
vibrato	0.2521	[0.2439, 0.2603]***	5.5 cents

Table 3.2: Mean absolute pitch error (MAPE) and 95% confidence intervals for each stimulus type and differences from results for the stable stimulus(***p < .001; **p < .01; *p < .05).

Factor	Test Results
The combined score	$F(30, 18541) = 54.4$ ***
Active engagement	$F(21, 18550) = 37.3$ ***
Perceptual abilities	$F(22, 18549) = 57.5$ ***
Musical training	$F(24, 18547) = 47.2$ ***
Singing ability	$F(20, 18551) = 69.8$ ***

Table 3.3: Influence of background factors on MAPE(***p < .001).

influence results. Likewise most musical training includes some ear training which would be beneficial for this experiment.

3.4.3 Other factors of influence for absolute pitch error

More than 90.77% of the mean absolute errors are less than 1 semitone. 6.3% of the mean absolute errors are larger than 3 semitones because the participants sang the previous pitch rather than the same pitch as the stimulus. Figure 3.5 shows the distribution of the 75 notes. Note numbers 1-24 are head stimuli, 25-48 are tail stimuli, 49-60 are the ramp stimuli, 61-63 are the constant pitch stimuli and 64-75 are vibrato notes.

R (R Development Core Team, 2008) and *lme4* (Bates et al., 2015) were used to perform a linear mixed-effects analysis of the relationship between factors of influence and $|e^{\mathrm{p}}|$. The factors stimulus type, main pitch, age, sex, order of stimuli, trial condition, repetition, duration of pitch deviation $\mathbf{d}$, extent of pitch deviation $\mathbf{p_D}$, observed duration and the four factors describing musical background were added cumulatively into the model, and a one-way ANOVA between the models with and without the factor tested whether the factor had a significant effect. Table 3.4 shows the p-value of ANOVA results after adding
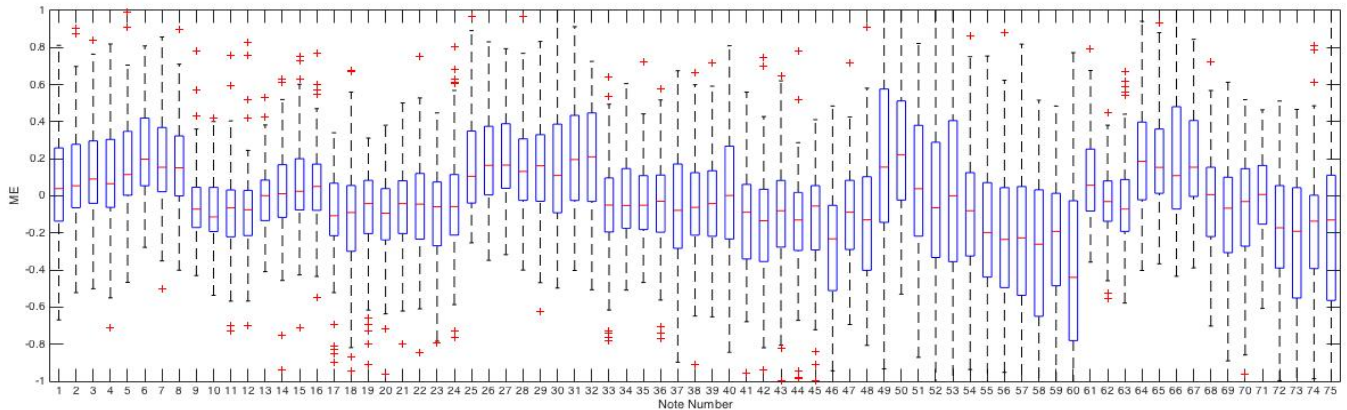
Figure 3.5: Mean pitch error by note number.

each factor in the order given.

A fixed model was created with the factors stimulus type, main pitch, repetition and trial condition, plus singer as a random effect. Visual inspection of residual plots did not reveal any obvious deviations from homoscedasticity or normality. The p-values were obtained by likelihood ratio tests of the full model with the effect in question, against the model without the effect in question (Winter, 2013).

The LME models gave different results for the background questionnaire factors than the one-way ANOVA; the effect size and significance are slightly different. According to the modelling results on $|e^p|$, significant effects were found for the factors stimulus type, main pitch p_m (effect size: -2.35 cents per octave), trial condition, repetition, musical background, duration of pitch deviation (effect size: 11.4 cents per second), direction of pitch deviation, magnitude of pitch deviation (effect size: 1.9 cents per semitone) and observed duration (effect size: -5.4 cents per second). The remaining factors (singer, age, sex and the order of stimuli) did not have any significant effect on $|e^p|$ in this model.

Contrary to the hypothesis, singing simultaneously (MAPE: 0.26; SD: 0.25) is 3.2 cents less accurate than the sequenced condition (MAPE: 0.23; SD: 0.21). Despite the large spread of results, the standard errors in the means are small and the difference is significant. Recall also that responses with $|e^p|$ over 2 semitones were excluded.

Other significant factors were repetition, where MAPE decreases 1.8 cents for each repetition (that is, participants improved with practice), and observed duration and main pitch, which although significant, had very small effect sizes for the range of values they took on.

Table 3.4: Significance and effect sizes for tested factors based on ANOVA results between pairs of linear mixed-effect models.

Factors	p-value	Effect size (cents)
Stimulus type	2.2e-16***	See Table 3.2
p_m	5.4e-7***	-0.19
Age	0.51	
Sex	0.56	
Order of stimuli	0.13	
Trial condition	2.2e-16***	3.2
Repetition	2.2e-16***	-1.8
Duration of transient d	2.2e-16***	11.4
$\text{sign}(p_D)$	5.1e-6***	0.8
$\text{abs}(p_D)$	8.3e-12***	1.9
Observed duration	3.3e-4***	-5.4
Active engagement	6.9e-2	
Perceptual abilities	0.04*	-0.3
Musical training	6.2e-5***	-0.5
Singing ability	8.2e-2	

3.4.4 Effect of pitch deviation on pitch error

The next step is to look at specific effects on the direction of pitch error, to test the hypothesis that asymmetric deviations from main pitch are likely to lead to errors in the direction of the deviation. For the *stable*, *head* and *tail* stimuli, a correlation analysis was conducted to examine the relationship between pitch deviation and MPE. The result was significant on MPE ($F_{(4, 12642)} = 8.4$, $p < .001$) and MAPE ($F_{(4, 12642)} = 8.2$, $p < .001$). A significant regression equation was found, with $R^2 = 2.5e - 3$, modelling pitch error as $e^P = 0.033 + 0.01 \times p_D$. Pitch error increased 1 cent for each semitone of p_D, a significant but small effect, as shown in Figure 3.6.

3.4.5 Duration of transient

As predicted, the duration d of the transient has a significant effect on MPE ($F_{(5, 18566)} = 51.4$, $p < .001$). For the *stable*, *head* and *tail* stimuli, the duration of transient influences MAPE ($F_{(2, 12644)} = 31.5$, $p < .001$), where stimuli with smaller transient lengths result
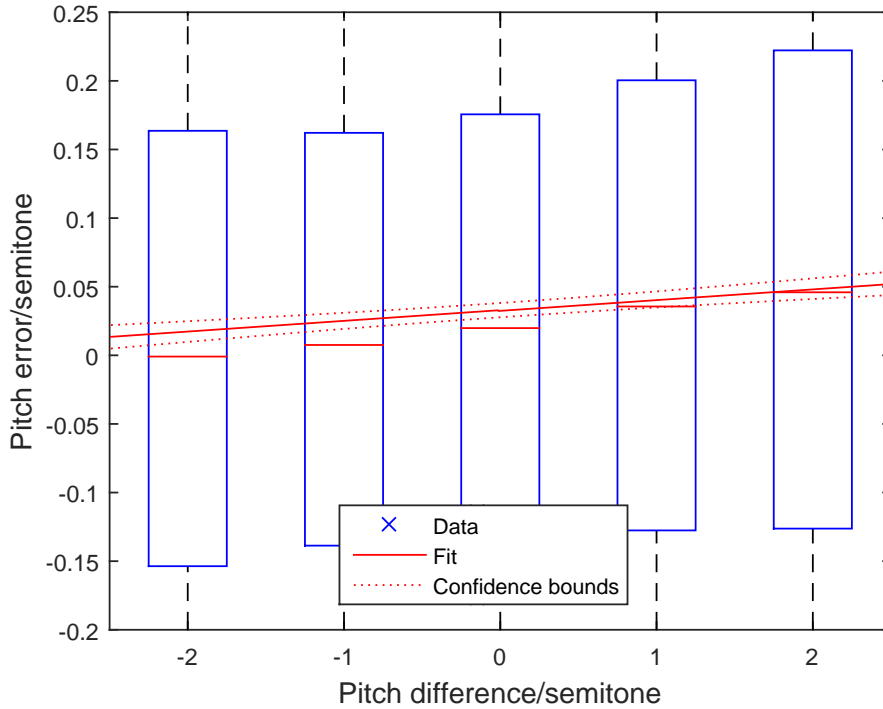
Figure 3.6: Boxplot of MPE for different p_D, showing median and interquartile range, regression line (red, solid) and 95% confidence bounds (red, dotted). The regression shows a small bias due to the positively skewed distribution of MPE.

in lower MAPE values. The regression equation is $\text{MAPE} = 0.33 + 0.23d$ with $R^2 = 0.208$. The MAPE increased by 23.2 cents for each second of transient. This matches the result from the linear mixed-effects model, where the effect size is 23.8 cents per second.

Based on the modelling results, transient length in responses was longer than in the corresponding stimuli. 74.2% of *head* and *tail* responses were found to have transient length longer than that of the stimulus. Stimulus transients are 0.1 or 0.2 seconds, but 65.5% of *head* and 72.0% of *tail* responses were found to have a transient longer than 0.2 seconds.

3.5 Modelling

In this section, the observed pitch trajectories were fitted to models defined by the stimulus type, to better understand how participants imitated the time-varying stimuli. The *head* and *tail* responses were modelled by a piecewise constant and quadratic function. Meanwhile the *stable* responses were modelled by a constant, the *ramp* by a linear fit, and the *vibrato* by a raised sinusoid.
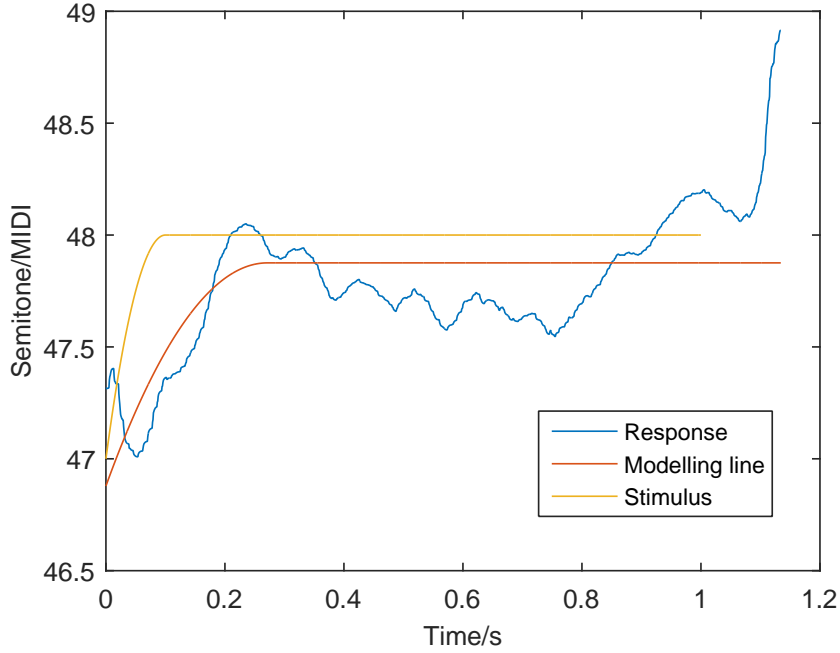
Figure 3.7: Example of modelling the response to a *head* stimulus with parameters $d = 0.1$, $p_D = -1$ and $p_m = 48$. The response model has $\hat{d} = 0.24$, $\hat{p_D} = -0.997$ and $\hat{p_m} = 47.87$. The forced fit to the stimulus model treats as noise response features such as the final rising intonation.

3.5.1 Initial and Final Transients

For the initial and end transients, two piecewise functions were set as the modelling shape. Given the break point, corresponding to the duration of the transient, the two parts can be estimated by regression. A grid search was performed on the break point, and the optimal parameters were selected according to the smallest mean square error.

Figure 3.7 shows the modelling results of an observed note (singer No. S031, trial 2, stimulus No. 19, in second repetition) for which the modelling parameters are: $\hat{d} = 0.24$, $\hat{p_D} = -0.997$ and $\hat{p_m} = 47.87$. The stimulus that was provided has the parameters: $d = 0.1$, $p_D = -1$ and $p_m = 48$. From modelling all the notes in this way, it was found that 78.87% of the notes have pitch difference less than 2 semitones. 22% of the transient portions are less than 0.2 seconds in duration.

Figure 3.8 is an example of modelling a *tail* stimulus (singer No. P05, trial 1, stimulus No. 32, in second repetition) where the modelling parameters are: $\hat{d} = 0.24$, $\hat{p_D} = 3.1125$, $\hat{p_m} = 60.0381$. The stimulus parameters were: $d = 0.2$, $p_D = 2$, $p_M = 60$. From modelling all the notes in this way, it was found that 73.97% of the $\hat{p_D}$ is less than 2 semitones, 43.26% of notes have $\hat{d}$ less than 0.2 seconds.

Figure 3.9 shows the distribution of $\hat{d}$ where most of the models do not have a change point. They are smooth or modelled by a quadratic equation. Although the stimuli that
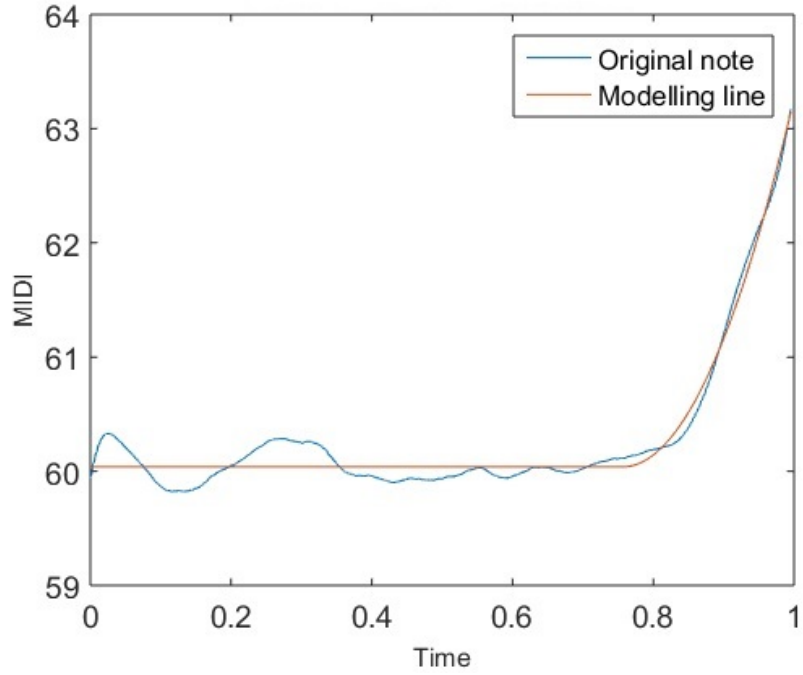
Figure 3.8: Example of a response to a *tail* stimulus and the best fitting model.

participants heard had a various change points, the responses were smooth and without a change point.

3.5.2 Ramp and Constant Pitch

The model $\hat{p_m}$ of a *ramp* response is the median of $p(t)$ for the middle 80% of the response duration. The linear regression to model the slope shape is given by Equation 3.3.
Figure 3.10 shows a modelling example (singer No. S020, trial 1, stimulus No. 60, in third repetition) for which $\hat{p_m} = 69.4530$, $\hat{p_D} = 2.1491$. The stimulus has $p_m = 70$ and $p_D = 2$

(a) *head* stimulus
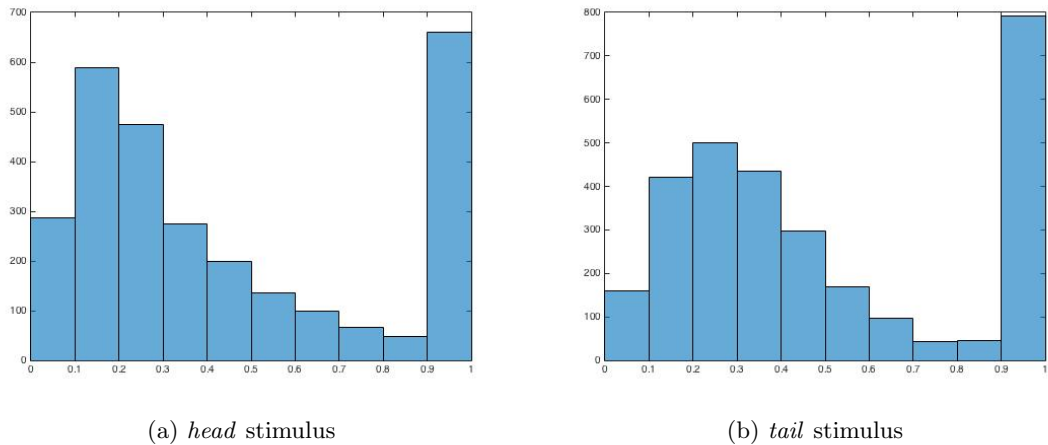
(b) *tail* stimulus

Figure 3.9: Histogram of $\hat{d}$ of *head* and *tail* stimulus.

66
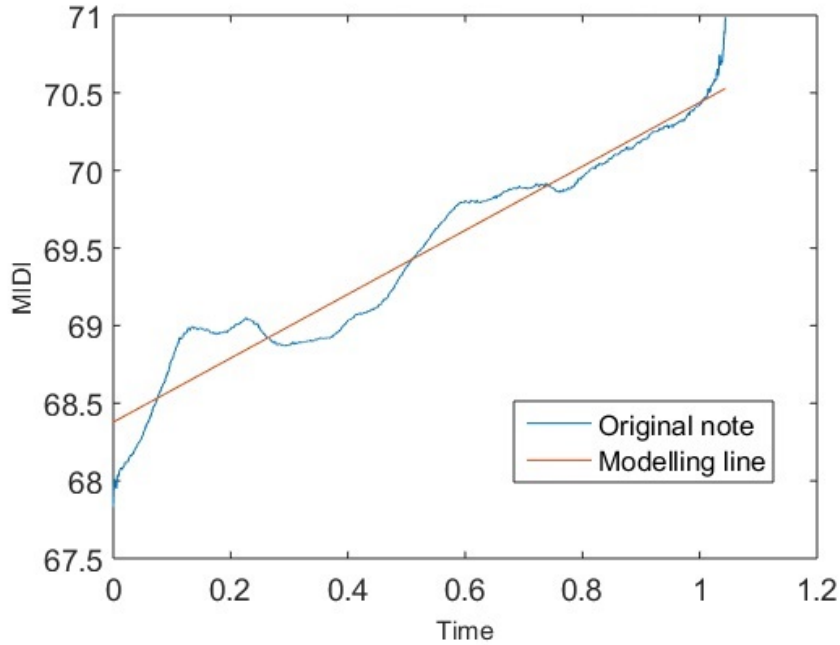
Figure 3.10: Example of a response to a *ramp* stimulus and the best fitting model.

semitone. 59.76% of the slope models have p_D smaller than 2 semitone.

The *stable* stimulus is modelled by a constant function whose parameter is calculated as the median of the pitch data, omitting the first 10% and last 10% of the data.

3.5.3 Vibrato

The *vibrato* responses were modelled with the MATLAB nlinfit function using Equation 3.4 and initialising the parameters with the parameters of the stimulus.

Vibrato notes were modelled by optimisation to fit Equation 3.4. This model needs three parameters p_m, p_D, and d. Figure 3.11 shows an example of vibrato modelling (singer No. S026, trial 1, stimulus No 66, in second repetition). The modelling results are: $\hat{d} = 0.2715$, $\hat{p_D} = -2.9690$, $\hat{p_m} = 60.0077$. The parameters of the stimulus are: $d = 0.5$, $p_D = 3$, $p_m = 60$.

Some of the *vibrato* models did not fit the stimulus very well because the singer attempted to sing a stable pitch rather than imitate the intonation trajectory.

3.6 Discussion

Since the target participants were non-poor singers, most participants imitated with small errors. 88.5% of responses were sung with intonation error less than half a semitone for the main pitch p_m. The responses are characterised far more by imprecision than inaccuracy.

67
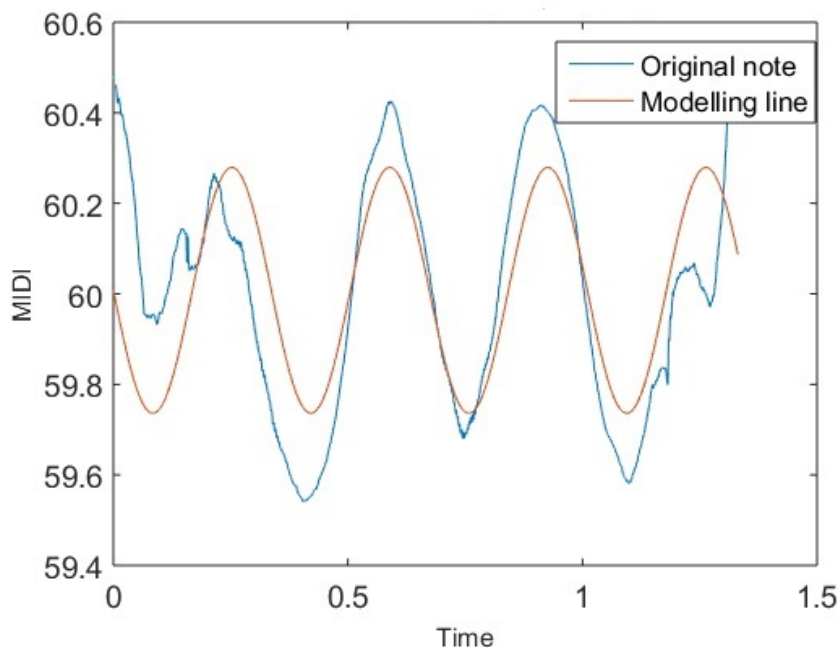
Figure 3.11: Example of a response to a *vibrato* stimulus and the best fitting model.

That is, there is very little systematic error in the results (MPE = 0.0123), whereas the individual responses exhibit much larger errors in median pitch (MAPE = 0.2441) and on a frame-wise level within notes ($\text{MAPE}_\text{f} = 0.3968$). The results for MAPE are within the range reported for non-poor singers attempting known melodies (19 cents by Mauch et al. (2014), 28 cents by Dai et al. (2015)), and thus are better explained by limitations in production and perception rather than by any particular difficulty of the experimental task.

The *stable* stimuli gave rise to the lowest pitch errors, although the *head* responses were not significantly different. The larger errors observed for the *tail*, *ramp* and *vibrato* stimuli could be due to a memory effect. These three stimulus types have in common that the pitch at the end of the stimulus differs from p_M. Thus the most recent pitch heard by the participant could distract them from the main target pitch. The *ramp* stimuli, having no constant or central pitch, was the most difficult to imitate, and resulted in the highest MAPE. The mean absolute error of *head*, *tail*, *ramp*, *stable* and *vibrato* shapes are 0.6165, 1.1358, 1.7938, 0.5720 and 0.7187 semitones respectively. The *ramp* stimulus has the largest modelling errors.

It was hypothesised that the simultaneous condition would be easier than the sequenced condition, as singing tends to be more accurate when accompanied by other singers or instruments. There are two possible reasons why this experiment might be exceptional. Firstly, in the sequenced condition, the time between stimulus and response was short

(1 second), so it would be unlikely that the participant would forget the reference pitch. Secondly, the stimulus varied more quickly than the auditory feedback loop, the time from perception to a change in production (around 100ms, Burnett et al. (1998)), could accommodate. Thus the simultaneous stimulus acts as a distractor rather than an aid. The results in Chapter 4 and 5 are consistent with this result.

Singing in practice requires staying in tune with other singers and instruments. If a singer takes their reference from notes with large pitch fluctuations, especially at their ends, this will adversely affect intonation.

3.7 Conclusion

A novel experiment was presented to test how singers respond to controlled stimuli containing time-varying pitches. 43 singers vocally imitated 75 instances of five stimulus types in two conditions. It was found that time-varying stimuli are more difficult to imitate than constant pitches, as measured by absolute pitch error. In particular, stimuli which end on a pitch other than the main pitch (*tail*, *ramp* and *vibrato* stimuli) had significantly higher absolute pitch errors than the *stable* stimuli, with effect sizes ranging from 15 cents (*ramp*) to 4.1 cents (*tail*).

Using a linear mixed-effects model, the results determined that the following factors influence absolute pitch error: stimulus type, main pitch, trial condition, repetition, duration of transient, direction and magnitude of pitch deviation, observed duration, and self-reported musical training and perceptual abilities. The remaining factors that were tested had no significant effect, including self-reported singing ability, contrary to other studies (Mauch et al., 2014).

Using one-way ANOVA and linear regression, a positive correlation was found between extent of pitch deviation (pitch difference, p_D) and pitch error. Although the effect size was small, it was significant and of similar order to the overall mean pitch error. Likewise, the duration d of the transient proportion of the stimulus correlated with absolute pitch error. Contrary to expectations, participants performed 3.2 cents worse in the condition when they sang simultaneously with the stimulus, although they also heard the stimulus between singing attempts, than in the sequenced condition.

Finally, parameters of the responses were extracted by a forced fit to a model of the stimulus type, in order to describe the observed pitch trajectories. The resulting parameters matched the stimuli more closely than the raw data did. The experiment led to a new research question, how singers respond when performing with a live partner (next chapter).

3.8 Data Availability

Annotated data and code to reproduce the results are available at: `https://code.soundsoftware.` `ac.uk/projects/stimulus-intonation/repository`. A tutorial video for participants is at: `https://www.youtube.com/watch?v=xadECsaglHk`.

Chapter 4

Duet Interaction Study

This chapter presents an experiment to investigate singing interaction by analysis of the factors influencing pitch accuracy of unaccompanied duet singers. Eight pairs of amateur singers sang two excerpts either in unison or two-part harmony. Different experimental conditions were applied, varying which singers could hear their partners. After semi-automatic pitch-tracking and manual checking, the pitch error and interval error were calculated, and the factors of influence were tested by using a one-way ANOVA. The results indicate that: 1) singing with the same vocal part is more accurate than singing with a different vocal part; 2) singing solo has less pitch error than singing with a partner; 3) pitch errors are correlated, as singers adjust their pitch to mitigate their partner's error and preserve harmonic intervals at the expense of melodic intervals and absolute pitch; 4) other factors influence the pitch accuracy, including: score pitch, score harmonic interval, score melodic interval, musical background, vocal part and individual differences.

4.1 Methodology

This section describes the hypotheses, the experimental design, musical material, participants and experimental procedure. For the experiment, two *singing conditions* are defined: the *unison condition*, where two singers sing the same vocal part, and the *duet condition*, where they sing different vocal parts. There are also four *listening conditions*. In the *solo* condition, the two singers cannot hear each other. The two *simplex* conditions are where only one singer can hear the other singer (in either direction). The singer who cannot hear her partner is called the *independent singer* while the singer who hears her partner is the *dependent singer*. The *duplex* condition is where both singers can hear each other. Note that according to these definitions, both singers are independent in the solo condition,

and both are dependent in the duplex condition. Singers can hear their own voice in all conditions.

4.1.1 Hypotheses

Based on previous research and musical experience, five hypotheses were formulated according to the expected observations when singers interact. The experimental method was designed to test these hypotheses and quantify the extent of the effects observed.

Hypothesis 1: The unison singing condition has less pitch error, melodic and harmonic interval error than the duet condition.

Participants sing the same pitch in the unison singing condition while they sing harmony in the duet condition. An observation from choral singing is that most singers, particularly those with less musical training, find it easier to sing their vocal part when others around them are singing the same part. Singing in harmony with different parts requires greater concentration, to avoid being distracted from one's own part.

Hypothesis 2: Independent singers have less pitch error than dependent singers.

Auditory feedback is essential for accurate intonation. As either noise (Mürbe et al., 2002) or simultaneously playing the target melody (Pfordresher and Brown, 2007; Dai and Dixon, 2016) reduces singers' accuracy, it is expected to observe this effect in both singing conditions.

Hypothesis 3: The duplex condition has less harmonic interval error than the solo condition.

When singers do not hear each other, their errors are independent as it is impossible for them to adjust their intervals according to their partner's intonation. When they can hear their partner, they adjust their pitch in order to reduce the harmonic interval error. Since most of the singers have choral experience, this hypothesis is based on the assumption that such singers are somewhat able to attune to other singers and sing harmoniously as a group, which is an important skill that is practised in their rehearsals (Bohrer, 2002).

Hypothesis 4: There is a positive correlation between the pitch error of the dependent singer and the independent singer in the simplex conditions.

The simplex condition allows for a one-way influence of the intonation of the independent singer upon the dependent singer. It is predicted that this influence will be seen not only in the magnitude of pitch errors (it is harder to sing well when distracted by an out of tune partner), but also in the direction of these errors (the dependent singer will adjust their pitch to reduce errors in vertical harmonies at the expense of absolute pitch error

and melodic interval error). Thus a significant correlation between the pitch errors of dependent and independent singers provides evidence of interaction. Although features of the score could explain correlation in the unison condition (e.g. where both singers compress leaps), this effect is predicted to hold also for the duet condition, where the score would not have a uniform effect on both singers.

Hypothesis 5: The within-note pitch variation of dependent singers is higher than that of independent singers.

The final hypothesis relates to the variation of pitch within each tone, which provides another view of interaction between singers. In the independent condition, any adjustment of pitch within a note arises from the singer's own feedback loop and involuntary noise in the vocal production system. In the dependent condition, there is also scope for intentional adjustment to improve harmonic intervals, as well as unintentional changes due to the distraction of hearing another singer.

4.1.2 Design

To test these hypotheses, a controlled experiment was designed and implemented involving two musical excerpts, two singing condition (unison and duet) and three types of listening conditions (solo, simplex, duplex), as listed in Table 4.1. Each trial involves two singers, denoted A and B. In the unison condition both singers sing the same vocal part (either the soprano or alto part). In the duet condition, singer A sings the soprano part and singer B the alto.

For the listening conditions, the solo condition acts as a control, where the two singers sing separately without hearing each other. In the two simplex conditions, only one singer can hear their partner, with the direction of auditory feedback being reversed between the two conditions. Finally in the duplex condition, both singers hear the voice of their partner. Except for the voice of their partner in certain listening conditions, there is no accompaniment during the experiment.

4.1.3 Musical Materials

The soprano and alto parts of two common choral pieces "Silent Night" (Gruber, c.1816) and "O Sacred Head, Now Wounded" (melody by Hassler, c.1601, harmonised by J.S. Bach, c.1729) were chosen as the experimental materials. These two pieces are examples of the traditional Western church choir repertoire with the former song being particularly well-known. The scores are based on the Open Hymnal Project 2008 `http://openhymnal.`

Singing Condition	Listening Condition	A sings	B sings	A hears B	B hears A
Unison	Solo	Soprano	Soprano	No	No
Unison	Simplex	Soprano	Soprano	Yes	No
Unison	Simplex	Soprano	Soprano	No	Yes
Unison	Duplex	Soprano	Soprano	Yes	Yes
Unison	Solo	Alto	Alto	No	No
Unison	Simplex	Alto	Alto	Yes	No
Unison	Simplex	Alto	Alto	No	Yes
Unison	Duplex	Alto	Alto	Yes	Yes
Duet	Solo	Soprano	Alto	No	No
Duet	Simplex	Soprano	Alto	Yes	No
Duet	Simplex	Soprano	Alto	No	Yes
Duet	Duplex	Soprano	Alto	Yes	Yes

Table 4.1: Experimental design for two singers A and B: singing and listening conditions.

org/. The pitch range is from A3 to E♭5 (soprano: B♭3 to E♭5; alto: A3 to G4) with various melodic and harmonic intervals up to a minor 7th. The second piece was shortened to its first 12 bars as shown in Figure 4.1 to match the lengths of the two pieces.

4.1.4 Participants

Although factors of age and sex affect pitch accuracy (Welch et al., 1997), they are not a target of this research. As the musical material consisted of soprano and alto parts, only female singers were recruited. Because this experiment required singers to maintain their own part while the other singer sang a different part, only participants who have choral experience were recruited. The singers who participated in the experiment came from the music society and *a capella* society of the university. In order to identify and exclude any poor singers (Pfordresher and Brown, 2007), the mean absolute melodic interval error (Equation 2.8) of each singer was calculated. It was planned to exclude any with an error greater than 0.5 semitones; no singer needed to be excluded.

16 female UK residents took part in this experiment, with an age range from 19 to 30 years old (mean: 23.1; median: 23.5; SD: 3.3); the sopranos have an age range from 19 to 27 years old (mean: 23.0; median: 24.0; SD: 3.0), while the altos have an age range from 19 to 30 years old (mean: 23.3; median: 22.5; SD: 3.4). All the participants were able to

Silent Night

John F. Young 1863

(a) A: Piece 1: Silent Night

O Sacred Head, Now Wounded

James W. Alexander, 1830

(b) B: Piece 2: O Sacred Head, Now Wounded

Figure 4.1: Musical material selected for the experiments.

sing the range from A3 to E♭5 naturally, and could sing both pieces independently.

Eight of the participants identified themselves as sopranos, the other eight as altos. The participants were selected according to their willingness to volunteer and singing competency. All of them are amateur singers with choir training weekly in the society. The grouping was chosen according to their time schedule and preference; most participants preferred the partners with whom they cooperated previously in the music society.

For testing the effect of training, all the participants completed a self-assessment questionnaire based on the Goldsmiths Musical Sophistication Index (Müllensiefen et al., 2014) (Appendix C). The proportion of singers having more than three years of choir experience is 54.2%; all have some instrumental training, and 50.0% of the participants have at least six years of formal training on musical instrument or voice.

4.1.5 Procedure

The study was conducted with the approval of the Queen Mary Ethics of Research Committee (approval number: QMREC1456).

The participants were grouped into eight pairs of singers, each consisting of one soprano (singer A) and one alto (singer B) by self-identification. Each pair participated in both the unison and duet singing conditions. For each singing condition, each singer sang the two pieces in each of the four listening conditions as a set of data, resulting in eight pairs of duet data sets, eight pairs of unison soprano and eight pairs of unison alto data sets collected in this experiment, each consisting of eight recordings. All 384 recordings were grouped and labelled with the pair number, music piece, experimental conditions and the singer's questionnaire results for analysis (2 singers × 2 pieces × 4 trials × (1 duet + 2 unison) × 8 groups=384).

Before the recording, the singers were given about half an hour to warm up and become familiar with the pieces. Participants practised their vocal parts with piano and their partners. The recording did not start until the participants could sing their vocal parts individually while their partner was singing the other part. At the beginning of each trial, participants heard instructions identifying the piece and condition and were given their own starting pitch repeated four times on a digital piano. For example, for the first trial in the first piece, both participants heard "Please sing the first piece alone", then the soprano heard the click and a piano play the pitch F4 four times, and simultaneously the alto heard the click and a piano play the pitch D4 four times simultaneously. During the trials, singers could hear a metronome and read the music score, but no further reference

pitch was provided, nor did the participants talk to each other until the trial was completed. Randomising the order of trials makes it difficult for participants to follow in the practical recording, therefore the trials were recorded in the same order with the same equipment (described below). To avoid any effect of vowel sound, and to assist annotation of note onset times, the participants were asked to sing the syllable /ta:/ rather than the lyrics. The participants could not see their partner during the trials. The total time of the experiment, including rehearsal, trials and questionnaire, was about one and a half hours. The experiment was performed in two acoustically isolated rooms at the author's university with facilities for multi-track recording (Morrell et al., 2011). The equipment included an SSL MADI-AX analogue to digital converter, two Shure SM58 microphones and sound isolating headphones (Beyer Dynamic DT100). All the tracks were controlled and recorded with the software Logic Pro 10. The metronome and the reference pitches were also given by Logic Pro. The two microphone signals and (for reference) the two headphone signals were recorded on four separate tracks with a sampling rate of 44100 Hz and stored in .wav format. The total latency of the system is 4.9 ms from microphone to headphone, where 3.3 ms is due to the processing time of Logic Pro and 1.6 ms (71/44100) due to the converter.

4.2   Data Analysis

This section describes the annotation procedure and the measurement of pitch error, melodic interval error, harmonic interval error and pitch variation. These metrics of accuracy are the dependent variables for hypothesis testing while test and listening conditions are the main independent variables.

4.2.1   Annotation

The software *Tony* (Mauch et al., 2015) was used to annotate the recording with fundamental frequencies as extracted by the pYIN algorithm (Mauch and Dixon, 2014), where the default sampling period for *Tony* is 5.8 ms (Section 2.5). The conversion of fundamental frequency to musical pitch p is calculated according to Equation 2.1. This scale is chosen such that its units are semitones, with integer values of p coinciding with MIDI pitch numbers, and reference pitch A4 ($p = 69$) tuned to 440 Hz. After automatic annotation, every single note was checked manually to make sure the tracking was consistent with the data and corrected if it was not. All the annotation work was done by the author independently. The annotation of all 384 files took over 31 hours, and resulted in a
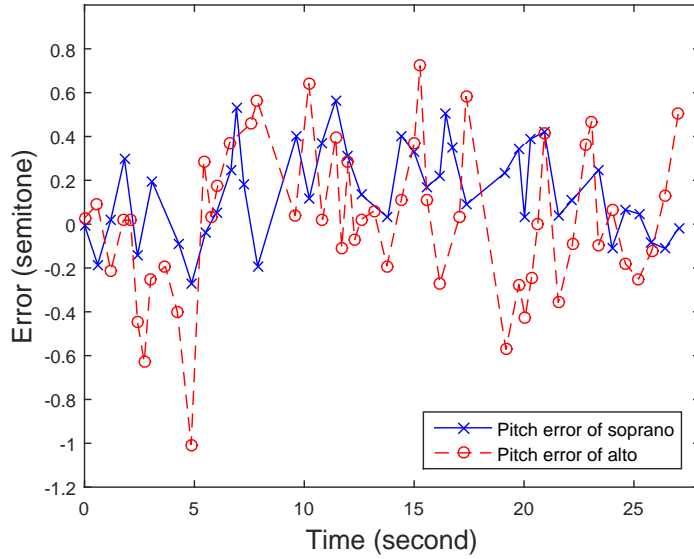
Figure 4.2: Example of pitch error for piece 2, duet singing condition, duplex listening condition, for one pair of singers.

database of 18176 annotated notes.

The information in the database includes: group number, singer number, singing condition, listening condition, piece number, note in trial, score onset position, score duration, score pitch, score interval, observed onset time, observed duration, observed pitch, pitch error, melodic interval error, harmonic interval error, anonymised participant details, and questionnaire scores. The pitch trajectory for each note was also stored. The data will be published for subsequent research (Section 4.7).

4.2.2 Metrics of Accuracy

The metrics of intonation accuracy are pitch error (the difference between the observed pitch and the score pitch), interval error (the difference between the observed interval between two adjacent pitches and the score interval) and pitch variation(the extent of pitch change within each single note). The definitions of pitch error and interval error are in Section 2.6.4, while pitch variability is defined in Section 2.3.3.

Pitch error measures the cumulative intonation error relative to the given starting tone. Figure 4.2 shows an example of pitch error for two singers in the duplex duet condition.

4.3 Statistical analysis

The calculation includes MAPE (Equation 2.4), MAMIE (Equation 2.8), MAHIE (Equation 2.9) and pitch variation (Equation 2.10) for each condition. In addition to the experi-

mental conditions, other possible factors were tested for their effect on singing intonation. Over all conditions, the singers had an MAPE of 36 cents, MAMIE of 24 cents and MAHIE of 41 cents. The MAPE was grouped according to different factors, and the grouped data was fitted separately into a one-way analysis of variance (ANOVA) model for testing the influence of each individual factor. The ANOVAs showed that the following factors influence the MAPE and MAMIE (MAPE and MAMIE have the same sample size and degrees of freedom): singing condition, listening condition, score pitch, score melodic interval, score harmonic interval, note duration, piece, vocal part, singer, age and musical background (Table 4.2). As MAHIE is defined as the harmonic interval difference, a pair of notes produce a harmonic interval error simultaneously, hence it cannot be grouped by some factors such as score pitch and vocal part. The statistical tests show that singing condition, listening condition, note number in trial, music piece and score harmonic interval have a significant effect on MAHIE.

Factor	MAPE	MAMIE	MAHIE
Singing condition	$F(1, 18174) = 70.8$ ***	$F(1, 18174) = 17.0$ ***	$F(1, 9086) = 316.7$ ***
Listening condition	$F(3, 18172) = 52.2$ ***	$F(3, 18172) = 41.0$ ***	$F(3, 9084) = 16.1$ ***
Note number in trial	$F(54, 18121) = 6.4$ ***	$F(54, 18121) = 15.2$ ***	$F(54, 9033) = 1.8$ ***
Score pitch	$F(15, 17552) = 22.3$ ***	$F(15, 17552) = 12.7$ ***	
Score melodic interval	$F(13, 18162) = 8.0$ ***	$F(13, 18162) = 90.6$ ***	
Score harmonic interval	$F(11, 18164) = 11.8$ ***	$F(11, 18164) = 13.5$ ***	$F(11, 9076) = 34.5$ ***
Score duration	$F(7, 18168) = 13.8$ ***	$F(7, 18168) = 94.5$ ***	
Piece	$F(1, 18174) = 102.7$ ***	$F(1, 18174) = 132.0$ ***	$F(1, 9086) = 121.5$ ***
Vocal part	$F(1, 18174) = 46.8$ ***	$F(1, 18174) = 58.8$ ***	
Age	$F(9, 18166) = 166.0$ ***	$F(9, 18166) = 59.4$ ***	
Musical background	$F(13, 18162) = 177.8$ ***	$F(13, 18162) = 77.6$ ***	

Table 4.2: Results of one-way ANOVA testing the MAPE, MAMIE, and MAHIE grouped by different factors.

In the data analysis section, single factors of influence were tested to investigate the hypotheses concerning intonation accuracy and pitch variation across the various experimental conditions.

4.3.1 Unison vs Duet Singing Condition

To test the first hypothesis, that the unison condition has lower pitch error and interval errors than the duet condition, a one-way ANOVA was conducted. For testing MAPE and MAMIE, only the data from dependent singers (those who can hear their partners)

was tested, which is one of the singers in the simplex listening condition and both singers in the duplex condition. Harmonic intervals involve both singers, so only the data from the duplex condition was tested for MAHIE. Results show a significant effect of singing condition on MAPE and MAHIE, but not for MAMIE (see Table 4.3). Post hoc comparisons using the Tukey honestly significant difference (HSD) test confirmed that MAPE and MAHIE were significantly lower for the unison condition than for the duet condition, for the unison condition, the score harmonic intervals are always 0 semitones. (Tukey HSD is a single-step multiple comparison procedure and statistical test, which can be used on raw data or in conjunction with an ANOVA (post-hoc analysis) to find means that are significantly different from each other.)

	Condition		Significance of Difference
	Unison	**Duet**	
MAPE	0.3518 ± 0.0057	0.4679 ± 0.0076	$F(1, 9086) = 149.38, p < .001$
MAMIE	0.2587 ± 0.0039	0.2637 ± 0.0052	$F(1, 9086) = 0.64, p = 0.42$
MAHIE	0.3447 ± 0.0060	0.5243 ± 0.0081	$F(1, 2270) = 262.23, p < .001$

Table 4.3: Results of one-way ANOVA testing the effect of singing condition on accuracy metrics.

The results confirmed the hypothesis for MAPE and MAHIE, but not for MAMIE. The reason for the higher MAPE in the duet condition (by 12 cents) may be due to the distraction of someone singing a different note, making it more difficult to sing one's own note than when the partner is singing the same note. For harmonic intervals, the duet condition has twelve different score intervals, while the unison condition has only one score interval, the unison interval. The various score intervals are more difficult to sing in tune, resulting in a higher MAHIE (by 38 cents) for the duet condition.

For MAMIE, there is no significant difference between the unison and duet conditions, so the results did not find any influence of singing condition on the tuning of melodic intervals. Melodic intervals are tuned from one's own previous note, while harmonic intervals are tuned between the singers; the other singers have no significant effect on the target melodic interval but significantly influence the harmonic intervals. The same argument, however, should also apply to pitch error, where a significant difference was observed. The relationship between the three error measures is complex, as any change in a single pitch will alter all measures. There is a tendency that when people sing different parts, their relative tuning to each other and absolute tuning to the initial reference suffer,

although their local melodic intervals do not, and it is easy to imagine that ideal singing with an imperfect partner would involve some compromise of all three error types.

4.3.2 Effect of Listening Condition

Hypotheses 2 and 3 predict that the solo listening condition has less pitch error but greater harmonic interval error than the duplex condition. ANOVA tests were conducted to test whether the four listening conditions have an influence on each measure of accuracy. Since the differences between listening conditions depend on whether singers can hear the voice of their partners, the data was separated from the simplex conditions into two cases: dependent singers and independent singers.

The ANOVA results showed that the effects of listening condition on MAPE, MAHIE and MAMIE were all significant: for MAPE, $F(3, 18172) = 52.16$, $p < .001$; for MAMIE, $F(3, 16956) = 38.77$, $p < .001$; and for MAHIE, $F(2, 9085) = 12.76$, $p < .001$. The ANOVA test tells whether there is an overall difference between groups, but it does not tell which specific groups differed. Post hoc comparisons using the Tukey HSD test were applied to find out which specific groups differed (Tables 4.4, 4.5 and 4.6).

	Significance of Difference			
	Solo	NS	***	***
		Simp. Indep.	***	***
			Simp. Dep.	***
				Duplex
MAPE	0.32	0.33	0.38	0.41

Table 4.4: Results of Tukey HSD test showing the effect of listening condition (solo, simplex independent, simplex dependent, duplex) on MAPE (***p<.001; **p<.01; *p<.05; NS: not significant).

	Significance of Difference		
	Solo	***	*
		Simplex	NS
			Duplex
MAHIE	0.45	0.39	0.41

Table 4.5: Results of Tukey HSD test showing the effect of listening condition (solo, simplex, duplex) on MAHIE (***p<.001; **p<.01; *p<.05; NS: not significant).

	Significance of Difference			
	Solo	**	***	***
		Simp. Indep.	***	***
			Simp. Dep.	NS
				Duplex
MAMIE	0.23	0.21	0.26	0.26

Table 4.6: Results of Tukey HSD test showing the effect of listening condition (solo, simplex independent, simplex dependent, duplex) on MAMIE (***p<.001; **p<.01; *p<.05; NS: not significant).

The results support hypothesis 2, as the MAPE of the simplex condition has 9 cents less pitch error than the duplex condition (Table 4.4). In general, participants have more pitch error when they can hear their partner singing than when they sing independently. This applies not only to the solo and duplex conditions, but also to the simplex conditions; in all cases, independent singers (solo and simplex independent) have significantly less MAPE than dependent singers (simplex dependent and duplex).

The results also show that the MAPE of dependent singers in the simplex condition is better than that in the duplex condition. This difference can be explained by considering that the partner of the dependent singer is an independent singer, while the partner of the duplex singer is a dependent singer. It was noted above that independent singers have lower MAPE than dependent singers, and accordingly their partners, who hear them, also sing with less pitch error.

The results for hypothesis 3 are shown in Table 4.5. In agreement with the hypothesis, the duplex condition has less harmonic interval error than the solo condition, even though the pitch error and melodic interval error are greater. For MAHIE, there is also a significant difference between solo and simplex conditions ($p < 0.001$) but not between the simplex and duplex conditions ($p > 0.05$).

As shown in Table 4.6, dependent singers in the simplex and duplex conditions have more MAMIE than independent singers ($p < 0.001$ in all four cases). These results have a similar pattern to those obtained for MAPE. An unexpected significant difference was also found between the two independent conditions (where the singer can not hear her partner). The effect size is small (2 cents), and can be explained as a learning effect, as the solo condition preceded the simplex conditions.

4.3.3   Correlation of Dependent and Independent Singers' Errors

The next step is to test hypothesis 4, whether there is a linear relationship between the pitch error (PE) of dependent and independent singers in the simplex condition. A linear regression was performed to model the pitch error of the dependent singer e_D^p as a function of the pitch error of the independent singer e_I^p (Figure 4.3), using the data from the duet condition only. A significant regression equation was found, $e_D^p = 0.02 + 0.91 \times e_I^p$ ($p < .001$), with $R^2 = 0.28$. The unison singing condition also exhibited a significant linear relationship, but with a smaller slope than in the duet condition.
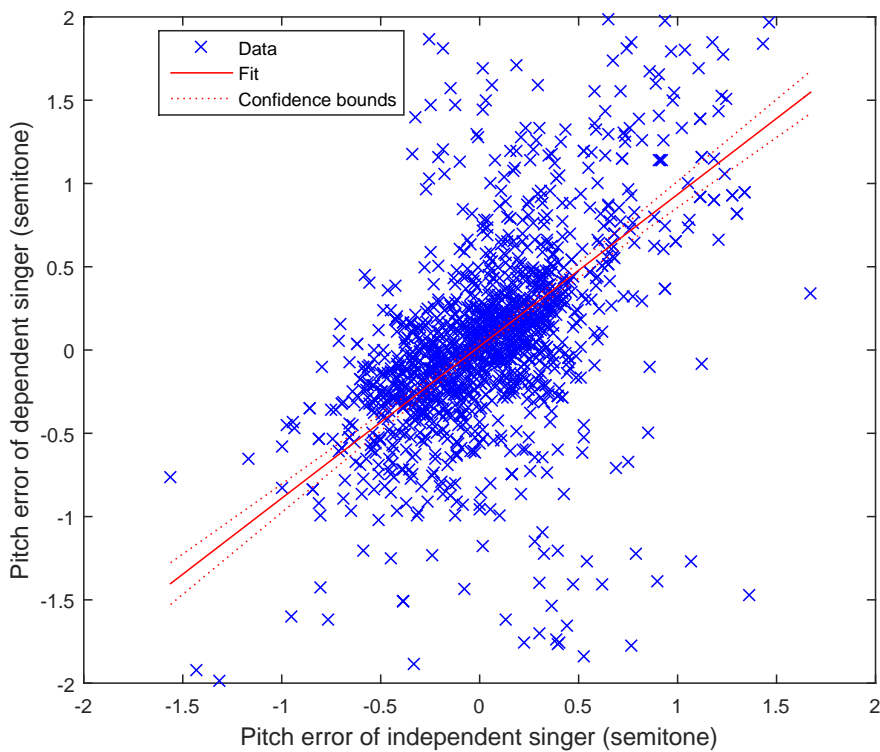
Figure 4.3: Scatter plot showing the correlation between independent and dependent singers' pitch error in the duet singing condition and simplex listening condition.

The melodic interval error (MIE) of dependent singers is also positively correlated to the MIE of independent singers ($r = 0.41$, $p < 0.001$) in the duet condition. The weak linear relationship is described by the following formula: $e_D^m = 0.005 + 0.59 \times e_I^m$, with $R^2 = 0.17$. There was also a significant but weak linear relationship between pitch variation of dependent singers and independent singers ($r = 0.12$, $p < 0.001$).

4.3.4 Pitch Variation within Notes

Hypothesis 5 concerns the pitch variation of dependent and independent singers. Pitch variation (Equation 2.10) does not show any significant effect of listening condition ($F(3, 17564) = 1.47, p = 0.22$). Likewise, an ANOVA applied to the two groups dependent singer and independent singer does not show a significant difference ($F(1, 17566) = 1.74, p = 0.19$). Thus the results fail to confirm the final hypothesis. The hypothesis had expected to find evidence of singers adjusting to their partner's pitch during a note. Some pairs of participants show a significant difference, where the pitch variation of dependent singers is higher than that of independent singers, as predicted, but this effect was not consistent across the whole dataset.

Moreover, the pitch variation in the unison condition (mean: 0.09; SD: 0.14) is lower than in the duet condition (mean: 0.11; SD: 0.16), with a statistically significant difference ($F(1, 17566) = 53.95, p < .001$). The pitch trajectories of the unison condition tend to be flatter in shape than those of the duet condition. There are a few factors that significantly influence pitch variation: the piece ($F(1, 17566) = 52.61, p < .001$), individual differences ($F(15, 17552) = 53.62, p < .001$), and score pitch ($F(15, 17552) = 20.6, p < .001$), where the high pitches (D5, E♭5) in particular exhibit greater variation. Thus pitch variation appears to reflect uncertainty of the singer in trying to reach the intended pitch, rather than deliberate adjustments to improve intonation.

4.3.5 Factors Based on the Score

The target pitch and its melodic and harmonic context are also expected to influence singing accuracy. These factors were tested with a series of ANOVAs. Score pitch ($F(15, 17552) = 22.23, p < .001$), score melodic interval ($F(13, 18162) = 7.99, p < .001$) and score harmonic interval ($F(11, 18164) = 11.8, p < .001$) all have a significant effect on MAPE. Likewise for MAMIE, score pitch ($F(15, 16346) = 10.88, p < .001$), score melodic interval ($F(13, 16946) = 89.02, p < .001$) and score harmonic interval ($F(11, 16948) = 13.3, p < .001$) all have a significant effect.

Although the score pitch has a significant effect on MAPE, the correlation between them does not show a linear tendency or a regular pattern. Higher pitch does not lead to a higher MAPE. After calculating the MAPE of different score pitches with the same note number, the most accurate pitch is C4 (0.2595 mean ±0.0089 confidence interval) while the least accurate pitches are A3 (0.5144±0.0229) and D♯4 (0.4524±0.0106).

Figure 4.4 shows the MAMIE for each score interval. The errors group into three clusters
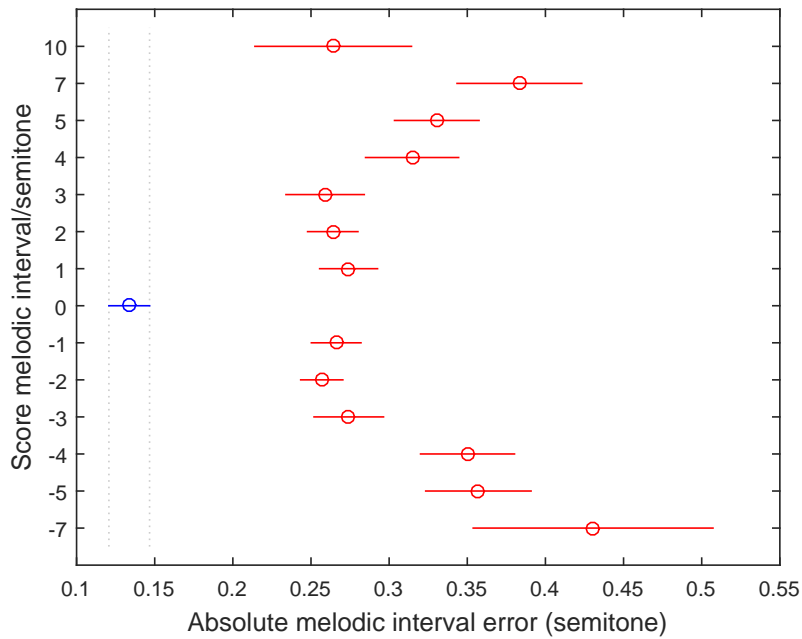
Figure 4.4: The mean estimates and the standard errors of absolute melodic interval error for each score melodic interval (significant differences from the unison interval are shown in red).

corresponding to (absolute) interval size. The unison interval has the smallest error, less than 15 cents, while intervals of one to three semitones have mean errors between 25 and 30 cents, and larger intervals have mean errors between 35 and 45 cents. All differences between clusters are significant, except for the ascending minor 7th (+10 semitone) interval, discussed below, and the ascending major third (+4), which lies on the border between the two clusters. Thus there is a general pattern of larger errors for larger intervals, with a small and non-significant tendency for descending intervals to have larger errors than their ascending counterparts. The ascending minor 7th interval is exceptional, being the largest interval, but having an error in the range of the smaller interval cluster. This interval only occurs twice, both times in the soprano part of the first piece. One explanation could be the lower error is because this melody (Silent Night) is particularly well-known.

The score harmonic interval has a significant effect on MAHIE ($F(11, 9076) = 34.48, p < .001$), as shown in Figure 4.5. Again the unison interval has the lowest error, and most score harmonic intervals have significant differences in MAHIE from the unison interval, except the major second and major sixth intervals. The least consonant intervals have the greatest error, with the minor second (mean:0.66; SD=0.98) and diminished fifth (mean:0.67; SD=0.79) having the largest MAHIE and also the largest spread of values.
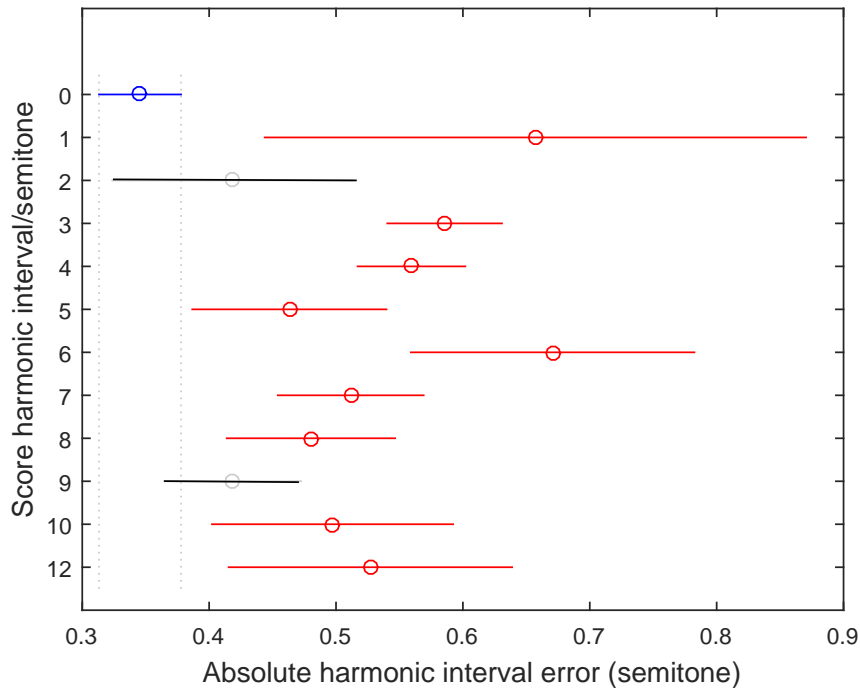
Figure 4.5: The mean estimates and the standard errors of absolute harmonic interval error for each score harmonic interval (significant differences from the unison interval are shown in red).

4.3.6 Vocal Part

The effect of vocal part (soprano, alto) on intonation accuracy was also investigated. Based on a one-way ANOVA, the vocal part has a statistically significant effect on MAPE ($F(1, 18174) = 46.78, p < .001$) and MAMIE ($F(1, 18174) = 58.76, p < .001$).

According to Section 4.3.1, the unison condition has less MAPE and MAMIE than the duet condition in general. However, there is an interaction with the factor of the vocal part. A two-way ANOVA was performed to examine the effect of singing condition and vocal part on MAPE. There is a significant interaction between the effects of vocal part and singing condition ($F(1, 18172) = 61.96, p < .001$). Simple main effects analysis (Table 4.7) showed that sopranos have significantly less MAPE than altos in the duet singing condition ($F(1, 6462) = 82.14, p < .001$) but there are no significant differences between vocal parts in the unison condition ($F(1, 11710) = 1.08, p = 0.30$). Further, the MAPE of the soprano part does not change significantly between the unison and duet conditions, but the alto part has a significantly larger MAPE in the duet condition as opposed to the unison condition. For MAMIE in both vocal parts, the duet condition has lower MAMIE than the unison condition, and in both conditions, the alto part has greater MAMIE than soprano.

Sopranos have less MAPE than the altos in the first piece but have more MAPE than altos

in the second piece. This result may be due to the the familiarity of the soprano part in Silent Night. Although the two vocal parts show different intonation accuracy according to the piece, the factor of vocal part does influence the pitch accuracy, and it also depends on the score and complex conditions.

	Unison	Duet	Significance: singing condition
MAPE Soprano	0.34	0.34	NS
MAPE Alto	0.34	0.44	***
Significance: vocal part	NS	***	
MAMIE Soprano	0.23	0.21	***
MAMIE Alto	0.26	0.25	**
Significance: vocal part	***	***	

Table 4.7: MAPE and MAMIE of soprano and alto in unison and duet singing conditions, and dependent listening conditions, showing the significance of differences between vocal parts and between singing conditions (***p<.001; **p<.01; *p<.05; NS: not significant).

4.3.7 Pitch Drift

Besides the previous factors, the note number in the trial also has a significant influence on MAPE ($F(54, 18121) = 6.44, p < .001$ in Table 4.2). Note number in trial is positively correlated with MAPE, which means that the absolute pitch error increases with time. The regression equation describing the relationship of note number in trial i and MAPE is: $MAPE = 0.235 + 0.002 \times i$, with $R^2 = 0.016$, $p < .001$. For each adjacent note, MAPE increases by 0.2 cents, resulting in about 10 cents of increase in MAPE from the beginning to the end of each trial.

The direction of the drift varies according to individual differences (Mauch et al., 2014; Dai et al., 2015); there was no overall trend to drift upwards or downwards. The magnitude of drift is similar to that found in a previous study (Mauch et al., 2014), where drift of 13.8 cents over 50 notes was found.

4.4 A Combined Model for Pitch Error

According to the results in Section 4.3, many single factors influence the pitch accuracy of individual singers and duets. In this section, the investigated factors were fitted to a
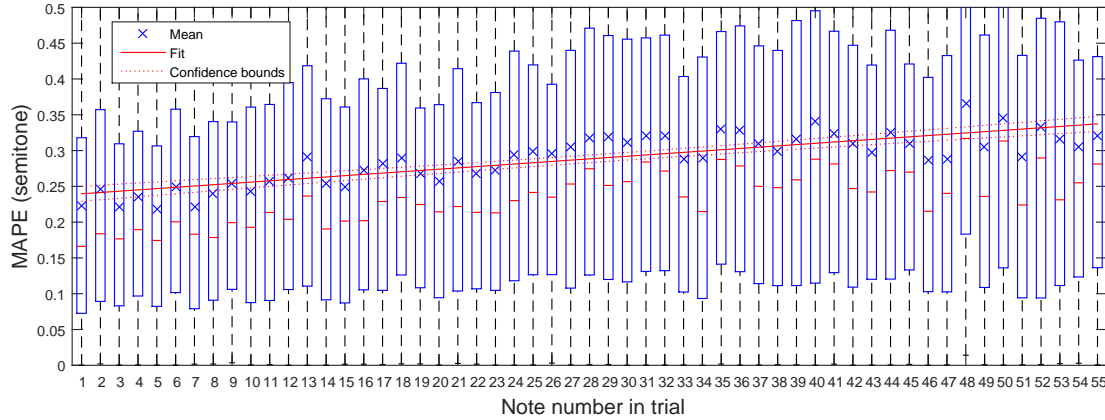
Figure 4.6: Box plot and linear fit showing that MAPE is positively correlated with note number in trial where X represent mean value of each group.

single linear mixed effects model for absolute pitch error, in order to understand how the factors influence pitch accuracy together.

The multiple factors were analysed using linear mixed-effects regression (LMER), using the `fitlme` function in Matlab and MAPE as the dependent variable. In contrast to the more traditional approach of data aggregation and repeated-measures ANOVA analysis, LMER controls for the variance associated with random factors without data aggregation. Before building the LMER model, the candidate factors were each tested with a one-dimensional linear regression. Some factors such as score pitch, score melodic interval, score harmonic interval, music background and note duration have a significant effect according to the ANOVA test, but their effect is not linear. Applying simple non-linear transformations to these variables does not change this fact: the effect of pitch and interval depends on the musical context, e.g. the tonality and the consonance or otherwise of the notes (see Figures 4.4 and 4.5); age has a limited range; musical background is sparse, dominated by individual factors; and duration is dominated by other score factors (the pitches of the longest and shortest notes). The factors which have a linear effect were added one by one into the LMER model and compared with the previous model, using 0.05 as the p-value threshold for rejecting insignificant factors.

Finally, the LMER model used singing condition, vocal part, listening condition and note number over time as the fixed effects. As a random effect, the model used the factor of the individual singer. Visual inspection of residual plots did not reveal any obvious deviations from normality. P-values were obtained by likelihood ratio tests of the full model with the effect in question against the model without the effect in question. Table 4.8 shows the resulting LMER model, where all the tested factors are significant.

Factor	Coeff.	SE	Significance
(Intercept)	0.0014	0.0500	NS
Note number in trial	0.0007	0.0002	**
Unison condition	-0.0378	0.0076	***
Simplex-dependent	0.0300	0.0103	**
Simplex-independent	0.0235	0.0103	**
Duplex	-0.0459	0.0100	***
Alto part	0.0528	0.0078	***

Table 4.8: A linear mixed-effects regression model for absolute pitch error, showing coefficient estimates (Coef.), standard errors (SE) and significance level of all predictors in the analysis (***p<.001; **p<.01; *p<.05; NS: not significant).

According to Table 4.8, the absolute pitch error can be described by the formula: absolute pitch error $= 0.0007 \times$ Note number in trial $- 0.0378 \times$ unison condition $+ 0.0300 \times$ simplex-dependent $+ 0.0235 \times$ simplex-independent $- 0.0459 \times$ duplex condition $+ 0.0528 \times$ alto condition $+ 0.0014$. The same process was followed for MAMIE and MAHIE. However, the models of MAMIE and MAHIE did not show a significant result in linear model.

To test for interaction effects between note number in trial, singing condition, listening condition, and vocal part, a factorial ANOVA was conducted. The interaction effect was significant for all the combinations of a pair of factors except the interaction effect of note number in trial $\times$ listening condition, and note number in trial $\times$ singing condition.

In the previous results (Section 4.3.1), the unison condition has less MAPE than the duet condition. However, it has the opposite tendency in the LMER model. When the LMER model was applied group by group, the effect size and tendency varied across groups (although most of them show a significant effect). For 3 of the groups, the duplex condition has a significant positive effect on MAPE, while 4 groups show a significant negative effect size, and one has no significant difference between conditions. To account for these group differences the model was refitted with random slopes for condition across groups. However, after refitting with random slopes, the listening conditions do not show any significant results in the LMER model. As noted in Chapter 2, previous studies do have conflicting results for listening condition, some of them observing a significant effect while others report non-significance. Although listening condition shows a significant effect on pitch accuracy, the tendency and effect size of listening condition might be due to individual differences.

	Effect size of duet condition	Significance
Group 1	0.4579	***
Group 2	-0.2893	***
Group 3	-0.0350	**
Group 4	-0.6215	***
Group 5	-0.1301	***
Group 6	-0.0091	NS
Group 7	0.1901	***
Group 8	0.0696	***

Table 4.9: The effect size and significance of the duet condition in the LMER model for each group (***p<.001; **p<.01; *p<.05; NS: not significant).

4.5 Discussion

It is evident that dependent singers adjusted their pitch under the influence of their partner's pitch. An important question to resolve is whether these adjustments were deliberate (e.g. to mitigate inaccuracies in their partner's singing), or inadvertent changes caused by the distraction of the partner's voice. Table 4.5 shows that the MAHIE in the simplex and duplex conditions is smaller than in the solo condition ($p < .001$). At the same time, singers who hear the voice of their partners (dependent singers) have higher MAPE and MAMIE than independent singers. Taken together, this supports the view that singers sacrifice some accuracy in singing their own part in order to harmonise (or sing in unison) better with their partner, and the accumulation of the intonation error may lead to pitch drift.

In this chapter, averages across singers (and their partners) were reported, not taking into account individual characteristics which may vary from pair to pair, for example the tendency of a singer to lead or follow, regardless of their partner's accuracy. One could characterise such tendencies by the extent of influence of the partner's singing, where a leader would be influenced less and a follower more by their partner's pitch. It is likely that such characteristics of interaction exist and influence the results, but the experimental design (each singer sings with a fixed partner) does not allow us to determine such cases unambiguously, as a singer's behaviour might arise in part from a reaction to their particular partner.

In a standard choral situation, multiple singers are assigned to each of several parts.

The experiment only considers the simpler case of two singers, and future works must use caution in extrapolating to the more general case. Conventionally, conductors group singers with the same vocal part together. The overall lower pitch error for the unison condition supports this practice, although the interaction with vocal part suggests that it might not be necessary for the sake of a dominant part such as soprano. Another choral practice supported by these results is to place weaker singers next to strong singers so that they can intentionally follow their pitch.

Although the participants in this study were selected as having vocal performance and choral experience, they are all amateur singers. They were given limited time to learn their parts (although one can assume that they already knew the melody of Silent Night), so some of the error might be due to lack of familiarity with the parts. Different results might have been obtained if the experiment had focused on professional singers, where the overall level of accuracy is likely to have been much higher.

4.6   Conclusions

This chapter presented an experiment investigating pitch accuracy and interaction in un-accompanied duet singing. 16 female participants sang two pieces of music in two singing conditions (unison and duet) and three types of listening condition (solo, simplex and duplex). The results indicated significant effects of the following factors on absolute pitch error: singing condition, listening condition, vocal part, and note number in trial, as well as score factors and individual factors of the singer. Likewise the melodic intervals and the harmonic intervals were affected by the same factors.

In terms of singing conditions, the unison condition has 12 cents less mean absolute pitch error and 38 cents less mean absolute harmonic interval error than the duet condition. This gives some measure of the additional difficulty of singing in harmony, and particularly of tuning non-unison intervals.

The general effect of singing with a partner is an increase in errors of individual pitches and intervals, but a reduction in the error of the interval between singers. That is, singers adjust their pitch to harmonise better with their partner, at the expense of continuity of tonal reference. Independent singers have 7 cents less pitch error than singers who can hear their partner.

The target harmonic interval has a significant effect on MAHIE, with dissonant intervals having the largest errors and the unison interval the smallest. For melodic intervals, the perfect fifth had the largest MAMIE, which is somewhat surprising considering the

previous result and the fact that it is a consonant interval. However it is one of the largest melodic intervals in the material (exceeded only by the two minor 7$^{\text{th}}$ leaps in the soprano part of Silent Night), and thus the results suggest the size of the interval to be a contributing factor in this case. I would expect consonance of intervals to play a smaller role for melodic intervals than harmonic intervals, since in the melodic case, the pitches do not sound simultaneously.

A positive correlation between the signed pitch errors of dependent singers and independent singers in the simplex condition was found. In other words, if one singer sings sharp, their partner is influenced to sing sharp as well. The correlation of pitch errors is again evidence of interaction, that singers adjust their pitch to improve harmonic intervals at the expense of melodic intervals and preservation of the tonal reference.

Analysis of the pitch trajectories within tones revealed greater stability of pitch in the unison condition than the duet condition, but not in independent singers over dependent singers. Although stability is correlated with singing accuracy, pitch variation is necessary if singers are to adjust dynamically to the pitch of an imperfect partner, which is what the research questions expected to find in the data. However, the results suggest that the observed pitch variation arises more from imprecision or uncertainty than deliberate adjustment. Further analysis of the pitch trajectories is addressed in Chapter 6.

There is considerable scope for further work on singing intonation and interaction, either by extending the analysis of the dataset, which is released as open data (Section 4.7), or by collecting further data for analysis. In particular, in order to move towards more typical musical settings, further studies need to investigate cases where there are multiple (more than two) singers per part, multiple parts, and instrumental accompaniment. In the following chapters, several quartets singing in an SATB setting are recorded and analysed.

4.7   Data availability

The code and the data needed to reproduce the results (note annotations, questionnaire results, score information) are available from:

https://code.soundsoftware.ac.uk/projects/pitch-accuracy-and-interaction-in-unacc ompanied-duet-singing/repository.

Chapter 5

SATB Study

In Chapter 3 and Chapter 4, various factors of the heard intonation were shown to affect the observed intonation accuracy. But how do singers respond when they cooperate with more than one vocal part? The aim of this chapter is to investigate interaction in four-part (SATB) singing from the point of view of pitch accuracy (intonation). In particular intonation accuracy of individual singers and collaborative ensembles were compared, which extends the previous experiments. 20 participants (five groups of four) sang two pieces of music in three different listening conditions: solo, with one vocal part missing and with all vocal parts. After semi-automatic pitch extraction and manual correction, the recordings were annotated and the pitch error, melodic interval error, harmonic interval error and note stability were calculated. Significant differences were observed between individual and interactional intonation, more specifically: 1) Singing without the bass part has less mean absolute pitch error than singing with all vocal parts; 2) Mean absolute melodic interval error increases when participants can hear the other parts; 3) Mean absolute harmonic interval error is higher in the one-way interaction condition than the two-way interaction condition; and 4) Singers produce more stable notes when singing solo than with their partners.

5.1 Research Questions

Unaccompanied ensemble singing is common in many musical cultures, yet it requires great skill for singers to listen to each other and adjust their pitch to stay in tune, while the quantitative analysis of the intonation needs further investigation.

This study of interactive intonation in unaccompanied SATB singing is driven by a number of research questions. The first aim is to determine whether singers rely on a particular

vocal part for intonation, which was tested by systematically isolating each vocalist so that the other singers cannot hear them. The bass part, which often contains the root notes of chords, is more important as a tonal reference (Terasawa, 2004), leading to **the first hypothesis: pitch error will be higher when the bass part is missing than when other voices are isolated.**

The second research question involves the effect of hearing other voices on intonation. The results from Chapters 3 and 4 suggest that singers are distracted by simultaneous sounds when they are singing, and they are less able to attend to their auditory feedback loop in order to sing accurately. This leads to **hypothesis 2, that the conditions in which singers hear no other voice will have less melodic interval error than in conditions when they hear other singers.**

This effect might be strengthened by conscious adjustment of singers to the other parts in order to improve the harmonic intervals. Thus as a corollary, this leads to the **third hypothesis, the harmonic interval error is lower when singers can hear each other than when they are isolated.**

An additional effect of interaction should be that singers adjust their pitch more during notes where they hear other singers (who might also be adjusting). **Thus the fourth hypothesis is that within-note variability in pitch will be higher (note stability will be lower) when singers hear each other than when they do not.**

5.2 Design & Implementation

A novel experiment was designed and implemented to test the hypotheses, which investigated the interaction between the four vocal parts. Three different listening conditions are defined, based on what the singer can hear as they sing.

5.2.1 Listening condition

In the *closed condition*, each singer hears no other voice than their own, thus they are effectively singing solo. That is, in the closed condition, each of Soprano, Alto, Tenor and Bass are isolated from each other and effectively singing solo. In the *partially-open condition* (or *partial condition* for short), the singer can only hear some, but not all of the other vocal parts. This is achieved by isolating one singer from the other three, and allowing acoustic feedback (via microphones and loudspeakers) in one direction only, either from the isolated singer to the other three singers (*one-to-three condition*), or from the three singers to the isolated one (*three-to-one condition*). Finally, in the *open condition*,

all singers can hear each other.

For example, one test condition is called the *soprano one-to-three condition*, where the soprano sings in a closed condition (i.e. she hears no other voice than her own), but all other parts - i.e. alto, tenor, bass - hear each other (including the soprano, whose voice is provided to the others via a loudspeaker). In such a case the isolated singer is called the *independent singer* as they are not able to react to the other vocal parts. As a further example, another test condition is the *soprano three-to-one condition*, where the soprano sings in an open condition (i.e. she hears all four parts (SATB)), but the alto, tenor and bass can only hear themselves (ATB). In cases where the singer can hear all (open condition) or some (partial condition) of the other voices, they are called a *dependent singer*. The recording of isolated *one-to-three condition* was used as the data for the closed condition.

For testing the partial condition, there are four pairs of test conditions corresponding to the vocal part that is isolated and the direction of feedback. For example, one test condition is called the *soprano isolated one-to-three condition*, where the soprano sings in a closed condition, but all other parts hear each other (the soprano's voice being provided to the others via a loudspeaker). In such a case the isolated singer is called the *independent singer* as they are not able to react to the other vocal parts to choose their tuning. In other cases the singer can hear all (open condition) or some (partial condition) of the other voices, and thus is called a *dependent singer*. Figure 5.1 gives an overview of the listening and test conditions where the arrows represent the direction of acoustic feedback.

5.2.2   Participants

20 adult amateur singers (10 male and 10 female) with choir experience volunteered to take part in the study. They came from the music society and *a capella* society of the university and a local choir. (There was also a pilot experiment involving four participants from our research group; this data is not used in this experiment.) The age range was from 20 to 55 years old (mean: 28.0, median: 26.5, SD: 7.8). Participants were compensated £10 for their participation. The participants were able to sing their parts comfortably and they were given the score and sample audio files at least 2 weeks before the experiment. Since training is a crucial factor for intonation accuracy, all the participants were given a questionnaire based on the Goldsmiths Musical Sophistication Index to test the effect of training (Müllensiefen et al., 2014). The participants had an average of 3.3 years of music lessons and 5.8 years of singing experience.
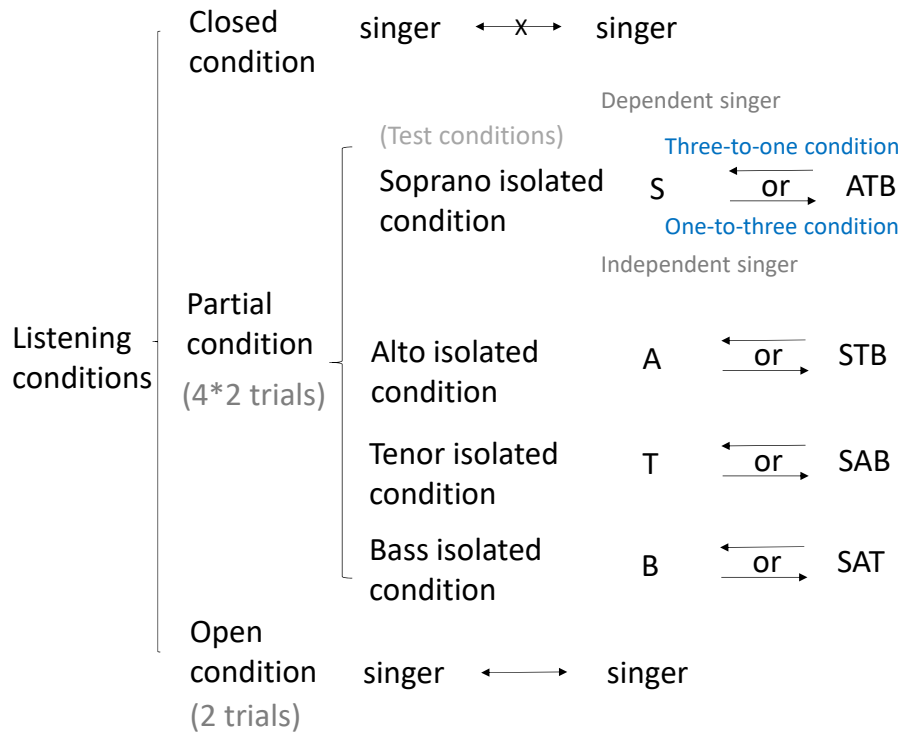
Figure 5.1: Listening and test conditions where the arrows present the direction of vocal accompaniment.

5.2.3 Materials

Two contrasting musical pieces were selected for this study: a Bach chorale, "Oh Thou, of God the Father" (BWV 164/6) and Leo Mathisen's jazz song "To be or not to be". Both pieces were chosen for their wide range of harmonic intervals: the first piece has 34 different harmonic intervals between parts and the second piece has 30 harmonic intervals. To control the duration of the experiment, the original score was shortened by deleting the repeat for both pieces. The tempo was also reduced from that specified in the score, in order to make the pieces easier to sing and compensate for the limited time that the singers had to learn the pieces. The resulting duration of the first piece is 76 seconds and the second song is 100 seconds. Links to the score and training materials can be found in Section 5.6 and Appendix B.

The equipment included an SSL MADI-AX converter, five cardioid microphones (Shure SM57) and four loudspeakers (Yamaha HS5). All the tracks were controlled and recorded by the software Logic Pro 10. The metronome and the four starting reference pitches were also given by Logic Pro. The total latency of the system is 4.9 ms (3.3 ms due to hardware and 1.6 ms from the software).

5.2.4 Procedure

The study was conducted with the approval of the Queen Mary Ethics of Research Committee (approval number: QMREC1560).

A pilot experiment with singers not involved in the study was performed to test the experimental setup and minimise potential problems such as bleed between microphones. Then the participants in the study were distributed into 5 groups according to their voice type (soprano, alto, tenor or bass), time availability and collaborative experience (the singers from the same music society were placed in the same group). Each group contained two female singers (soprano and alto) and two male singers (tenor and bass). Each participant had at least two hours practice before the recording, sometimes on separate days. They were informed about the goal of the study, to investigate interactive intonation in SATB singing, and they were asked to sing their best in all circumstances.

For each trial, the singers were played their starting notes before commencing the trial, and a metronome accompanied the singing to ensure that the same tempo was used by all groups.

Each piece was sung 10 times by each group. The first and the last trial were recorded in the open condition. The partial and closed condition trials, consisting of 8 test conditions, 4 (isolated voice) $\times$ 2 (direction of feedback), were recorded in between. The order of isolated conditions was randomly chosen to control for any learning effect. For each isolated condition, the three-to-one condition always preceded the one-to-three condition. The performance of isolated singers in the one-to-three conditions was used as the data for the closed condition.

The singers were recorded in two acoustically isolated rooms. For the partial and closed conditions, the isolated singers were recorded in a separate room from the other three singers. Loudspeakers in each room provided acoustic feedback according to the test condition. There was no visual contact between singers in different rooms. With the exception of warm-up and rehearsal, but including all the trials and the questionnaire, the total duration of the experiment for each group was about one hour and a half.

5.2.5 Annotation

The experimental data comprises 5 (groups) $\times$ 4 (singers) $\times$ 2 (pieces) $\times$ 10 (trials) = 400 audio files, each containing 65 to 116 notes. The software *Tony* (Mauch et al., 2015) was chosen as the annotation tool (Section 2.5).

For each audio file, two `.csv` files were exported, one containing the note-level information

(for calculating pitch and interval errors) and the other containing the pitch trajectories (for calculating pitch variability). All the intonations were relative to twelve-tone equal temperament, expressed in semitones according to MIDI standard pitch numbering. It took about 67 hours to manually check and correct the 400 files, resulting in 49200 annotated single notes, to which information was added on the singer (anonymous), score notes and metrics of accuracy. All the data are available in Section 5.6.

5.3 Data Analysis

The primary aim of this study was to test experimentally whether, and under what conditions, interaction is beneficial or detrimental to SATB intonation accuracy. The intonation accuracy of individuals was tested by pitch error (section 5.3.1), melodic interval error (section 5.3.2) and note stability (section 5.3.4); and the intonation of pairs of singers was tested by harmonic interval error (section 5.3.3). In order to avoid biasing mean errors by outliers, where a participant sang a wrong note rather than an out-of-tune attempt at the correct pitch, all the tests exclude notes with pitch error or interval error larger in magnitude than one semitone. 88.5% of observed notes had an absolute pitch error less than one semitone.

5.3.1 Pitch Error

The first task is to investigate whether the ensemble depends on a certain vocal part to tune their pitch. After excluding the notes which have an absolute pitch error larger than one semitone, most of the observed notes are relatively accurate (mean: 0.25 semitones; median: 0.26; SD: 0.07).

The pitch error was computed for the three non-isolated singers in each three-to-one condition and open condition, and results were analysed by test condition. The MAPE was computed as an average across the three non-isolated singers and the five groups. For example, in the soprano isolated three-to-one condition, the average of the pitch errors of alto, tenor, and bass parts were calculated from each group to give the resulting MAPE. Then these results were compared with the performance of the same three singers in the open conditions.

A correlated samples analysis of variance (ANOVA) showed a significant difference in MAPE between three-to-one and open conditions ($F(1,21625)=13$, $p<.001$). The MAPE of the three-to-one condition is less than the MAPE of the open condition. Then separate ANOVAs were applied for each isolated voice type (Table 5.1), and found that the results

vary across test conditions. The bass and tenor isolated three-to-one conditions both showed significant differences, while the results for the other two voice types were not significant.

Test condition	Partial vs open condition
Soprano isolated	$F(1,9391)=2.86$, $p=0.09$
Alto isolated	$F(1,9614)=0.61$, $p=0.11$
Tenor isolated	$F(1,9742)=5.07$, $p=0.02$*
Bass isolated	$F(1,10223)=14.39$, $p<.001$***

Table 5.1: Results of correlated samples ANOVAs for three-to-one and open listening conditions (***$p<.001$; **$p<.01$; *$p<.05$).

These results suggest that the bass part is the most influential vocal part in all observed groups. However, the direction of influence is the opposite of that hypothesised: removing the bass vocal part from the ensemble reduces the observed pitch error on average.

The next ANOVA shows that the MAPE is significantly different between the test conditions in the three-to-one listening condition ($F(3,12948)=28.67$, $p<.001$). Table 5.2 shows the 95% confidence intervals, which demonstrate that the bass and tenor isolated conditions are significantly different from all other three-to-one conditions. The bass isolated condition has 4 cents MAPE less than soprano and alto isolated conditions, and 2 cents MAPE smaller than the tenor isolated condition.

Test condition	MAPE	Confidence interval
Soprano isolated	0.2484	[0.2420, 0.2548]
Alto isolated	0.2483	[0.2422, 0.2545]
Tenor isolated	0.2328	[0.2271, 0.2385]
Bass isolated	0.2082	[0.2028, 0.2135]

Table 5.2: Mean absolute pitch error (MAPE) and 95% confidence intervals for three-to-one test conditions, for all non-isolated singers and all groups.

These results contradict hypothesis one: when singers do not hear the bass part, they sing more accurately on average, as shown by comparisons within the three-to-one conditions and between the three-to-one and open conditions.

5.3.2 Melodic Interval Error

To test the influence of interaction on adjacent notes within a voice (hypothesis two), melodic interval error was calculated. 91.9% of the note pairs have a melodic interval error smaller than one semitone (mean:0.21; median:0.21; SD:0.07).

A correlated-samples ANOVA was applied to test the effect of listening condition on MAMIE. The MAMIE is significantly different across listening conditions ($F(2,18333)=27.96$, $p<.001$). The listening condition of singing without hearing any partners (closed) has smaller MAMIE than the listening conditions with partners (partial and open). Table 5.3 shows the mean and confidence intervals for the three listening conditions where the closed listening condition has 3 cents smaller MAMIE than the open listening condition.

Listening condition	MAMIE	Confidence interval
Closed condition	0.1874	[0.1828, 0.1919]
Partial condition	0.2001	[0.1953, 0.2049]
Open condition	0.2138	[0.2102, 0.2174]

Table 5.3: Mean absolute melodic interval error (MAMIE) and 95% confidence intervals for each listening condition.

The acoustic feedback from other vocal parts increases MAMIE, which concurs with findings from previous chapters and the literature (Mürbe et al., 2002) and supports hypothesis two. The accompaniment from other vocal parts may mask the singer's own voice or distract the singer's attention from their own intonation. Alternatively, the increase in melodic interval error could be a side effect of deliberate adjustment of intonation to reduce harmonic interval error.

5.3.3 Harmonic Interval Error

Beside the intonation accuracy of individual singers, the accuracy of pairs of singers was also tested. There are four individual singers and up to six harmonic intervals simultaneously present at any point in time. All the harmonic intervals were observed under two circumstances: *one-way interaction* and *two-way interaction*.

In the partial conditions, some of the communication is only in one direction, so that any deliberate adjustment in harmonic interval must be attributed to the singer who can hear their partner. In this case, it is a one-way interaction. In the open conditions, both singers in a pair are able to adjust to each other, creating a two-way interaction. Taking soprano isolated conditions as an example, the harmonic intervals involving soprano are

one-way interactions, and the harmonic intervals between alto, tenor and bass are two-way interactions (Figure 5.2).
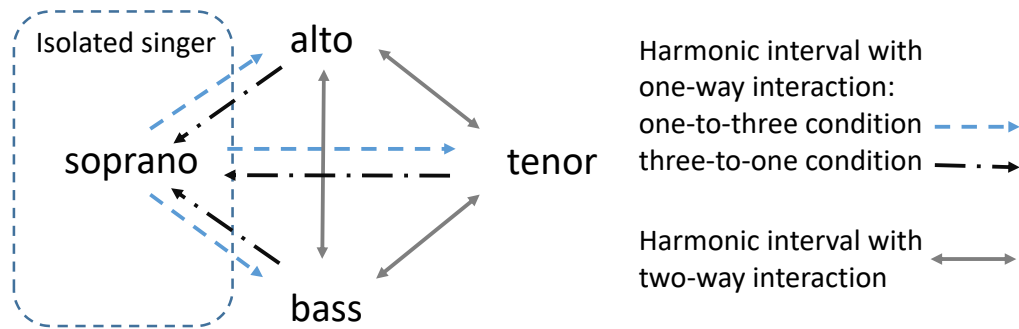

Figure 5.2: Interaction in the soprano isolated conditions.

By comparing the MAHIE for two-way interactions with those for one-way interactions in the three-to-one test conditions, the results show MAHIE is significantly smaller for the two-way interactions than for one-way interactions (F(1,23659)=10.94, p<.001). This supports the third hypothesis, and indicates that acoustic feedback helps singers to interactively tune harmonic intervals.

However, no significant difference was found between MAHIE for different directions of intonation, that is the three-to-one condition versus the one-to-three condition (F(1,23524)=0.39, p=0.53). When one side of interactive intonation is without acoustic feedback, the direction of the feedback does not appear to influence the harmonic interval.

5.3.4 Note Stability

The note stability is measured by its converse, note variability (Equation 2.10). The acoustic feedback of other singers not only has an influence on intonation accuracy (section 5.3.1 and 5.3.2) but also has an influence on note variability.

The note variability in the closed condition is significantly different from that in the partial and open conditions (F(1,23659)=41.23, p<.001), but no significant difference was found between the partial and open conditions (F(1,22514)=1.37, p=0.24). Note trajectories become less stable when singers can hear other singers in addition to their own voice, which is further evidence of interaction in intonation. This agrees with results from Chapter 3, as well as previous studies, which show that singers perform worse when singing with an unstable reference pitch (Pfordresher and Brown, 2007).

Moreover, the note variability is weakly positively correlated to the MAPE of individual notes (r=0.18, p<.001). Although the correlations to singer (r=0.01, p=0.01) and training (r=0.08, p<.001) are statistically significant, the effect sizes are negligible.

The fourth hypothesis has been tested, and the results confirm that there is a relationship between the listening condition and note stability. This complements results from other research which assert that note stability of individual singers depends on emotional expression (Fyk, 1995; Sundberg et al., 2013). Other possible relationships, such as a connection between musical training and note stability, were not supported by the experimental results.

5.4 Discussion

This study tested four hypotheses using various metrics of singing accuracy and statistical tests. In each case, significant results were found. In three of the four cases, the results supported the hypotheses, however for the first hypothesis, the direction of the observed effect was the opposite of what was predicted.

Participants noted that the bass part (male singer) is the most difficult vocal part to recruit. It is possible that this leads to a lower average standard among bass singers. A comparison of pitch error by vocal type reveals that the bass vocal part has a larger MAPE than the other vocal parts. This may be the cause of the unexpected result for the bass isolated condition: i.e. because the bass voice had greater pitch error, other parts which tuned to the bass also increased their pitch error.

As found in Chapter 4, the factor of interaction, which occurs when singers can hear each other, increases the pitch error of the individual singers but decreases the harmonic interval error between the singers. Although these results may appear to be contradictory, this can occur when melodic interval errors accumulate, and the sung pitches drift away from the initial tonal reference, as has been demonstrated by Howard (Howard, 2007c).

Many factors of influence have been researched which are crucial for singing, such as age and sex (boys are more likely to sing out of tune than girls), and individual differences (Welch et al., 1997). The analysis of singer-related factors and results from the questionnaire is deferred to the next chapter, where pitch trajectories from this experiment are investigated.

5.5 Conclusions

For analysis of the effect of interaction on intonation in unaccompanied SATB singing, a novel experiment was designed, which tested the intonation accuracy of five groups of singers in a series of test and listening conditions. The results confirm that interaction ex-

ists between singers and influences their intonation, and that intonation accuracy depends on which other singers each individual singer can hear.

In particular, it was observed that the three-to-one bass isolated test condition had a significantly lower MAPE compared with other three-to-one conditions, and compared with the open condition. In other words, singers were more accurate when they could not hear the bass. This surprising result might be due to the fact that the bass singers were less accurate on average than other singers in this experiment.

The results also show increases in pitch error and melodic interval error when singers could hear each other. The closed condition had the smallest MAMIE, while the open condition had the largest. At the same time, acoustic feedback decreased the harmonic interval error, while the direction of the feedback did not influence the harmonic interval error.

Interaction also has the effect of reducing the note stability, or increasing its variability. Pitch within a note varies more when singers hear each other, as one might expect if the singers are adjusting their intonation to be in tune with each other.

In conclusion, this experiment addresses a gap in singing intonation studies, by investigating the effects of interaction between singers. The results show that interaction significantly influences the pitch accuracy, leading to increases in the pitch error, melodic interval error, and note stability but a decrease in the harmonic interval error.

5.6 Data Availability

Annotated experimental data scores and code to reproduce the results are available at: https://code.soundsoftware.ac.uk/projects/analysis-of-interactive-intonation- in-unaccompanied-satb-ensembles/repository.

Chapter 6

Intonation Trajectories and Patterns of Vocal Notes

The previous chapters investigated the effects of interaction between singers on intonation accuracy via three coherent experiments. In Chapter 3, transient parts were observed and modelled, but the relationship between note trajectories and factors of influence need further research. Chapters 4 and 5 have investigated aspects of singing performance such as intonation accuracy and pitch drift, treating pitch as fixed within notes, while the pitch trajectory within notes has hardly been investigated. The aim of this analysis is to study pitch variation within vocal notes and ascertain what factors influence the various parts of a note. The data from Chapter 5 is used of five SATB quartets singing two pieces of music in three different listening conditions, according to whether they can hear the other participants or not. After analysing all of the individual notes and extracting pitch over time, the following regularities were observed: 1) There are transient parts of approximately 120 ms duration at both the beginning and end of a note, where the pitch varies rapidly; 2) The shapes of transient parts differ significantly according to the adjacent pitch, although all singers tend to have a descending transient at the end of a note; 3) The trajectory shapes of female singers are different from those of male singers at the beginnings of notes; 4) Between vocal parts, there is a tendency to expand harmonic intervals (by about 8 cents between adjacent voices).

6.1   Research Questions and Methodology

This section describes the exploratory research questions explored in this chapter. The experimental design, musical material, participants and experimental procedure are de-

scribed in detail in Section 5.2. Links to the data and score information can be found in Section 6.5.

This study of pitch trajectories in unaccompanied SATB singing is driven by a number of research questions. The first question is whether there are patterns or regularities in the pitch trajectories of individual notes. The aim is to find common trends in the note trajectories, with differences due to context and experimental conditions. The second question is how to characterise the trajectories in terms of the time required for the singer to reach the target pitch. The third question is what factors influence the tendencies of the transient part. The note trajectories might show significant differences due to contexts, such as when singing after a higher pitch or a lower pitch. Another research target is to determine whether pitch trajectories differ by vocal part or sex. In Chapter 5 significant differences between vocal parts in terms of pitch error were observed. The final question is whether the listening condition affects note trajectories. That is, do the shapes of vocal notes differ depending on whether the participants can hear other vocal parts or not?

6.2 Results

This section presents observed patterns in the shapes of note trajectories and investigates differences due to vocal part, sex, adjacent pitch and listening conditions, modelling the trajectories according to the shape of transient parts and classifying them into four categories.

Based on the metronome tempo, the expected duration of notes ranges from 0.25 to 5.50 seconds (mean 0.86, median 0.75), while the observed note duration is from 0.01 seconds to 5.10 seconds (mean 0.69, median 0.62). Any notes which had a duration shorter than 0.15 seconds (4.1%) or MAPE larger than one semitone (11.5%) were excluded.

6.2.1 The shape of note trajectories

To observe regularities in note trajectories across differing note durations, I truncate the time series and only consider the initial and final segments of each note. Taking the first (respectively last) 0.4 seconds of each note, excluding notes with a duration less than 0.55 seconds to avoid artefacts due to the transient at the other end of the note, results in the trajectories shown in Figures 6.1 and 6.2. It can be seen that the first 0.12 and last 0.12 seconds of each note have the most pitch variance.

The appearance of note trajectories is significantly different between singers who have different degrees of musical training. For the trained singers, the note trajectories are
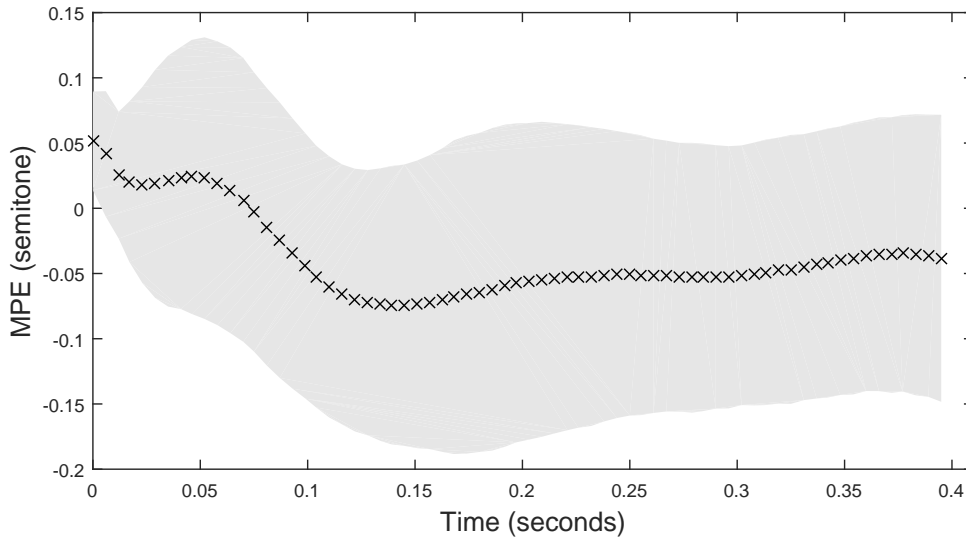
Figure 6.1: Mean pitch error for the initial 0.4 seconds of each note where the grey area shows the standard deviation.

smoother, and the two transient parts have a clear direction. For singers with less training, their note trajectories tend to be uneven and have a less common shape in their beginnings and endings.

The average results hide differences in the proportion and direction of transients which arise due to various influencing factors, for example individual difference, score pitch and vocal part. In Figure 6.1, the first turning point at 0.02 seconds may be an artefact of the averaging of different pitch trajectory shapes. There are several possible factors that might influence trajectory shapes, such as the pitch of the surrounding notes, vocal part, sex and listening condition, which are investigated in the following sections.

6.2.2 Adjacent pitch

The previous section observed large pitch fluctuations at each end of the note. To test whether these fluctuations are influenced by adjacent pitches in the score, the data for each end of a note was separated into two situations, based on whether the previous (respectively next) pitch is lower or higher than the current pitch. Repeated pitches are ignored. In both cases (previous and next note), an analysis of variance (ANOVA) confirms that the pitch error in relative time is significantly different based on whether the adjacent pitch is higher or lower. Figure 6.3 shows that singers tend to overshoot the target pitch and then adjust downward after singing a lower pitch, while after a higher pitch they reach the target almost immediately. The steady state pitch is 1 cent sharper when coming from a lower pitch than when the previous pitch is higher ($F(1, 38) = 77.97, p < 0.001$).
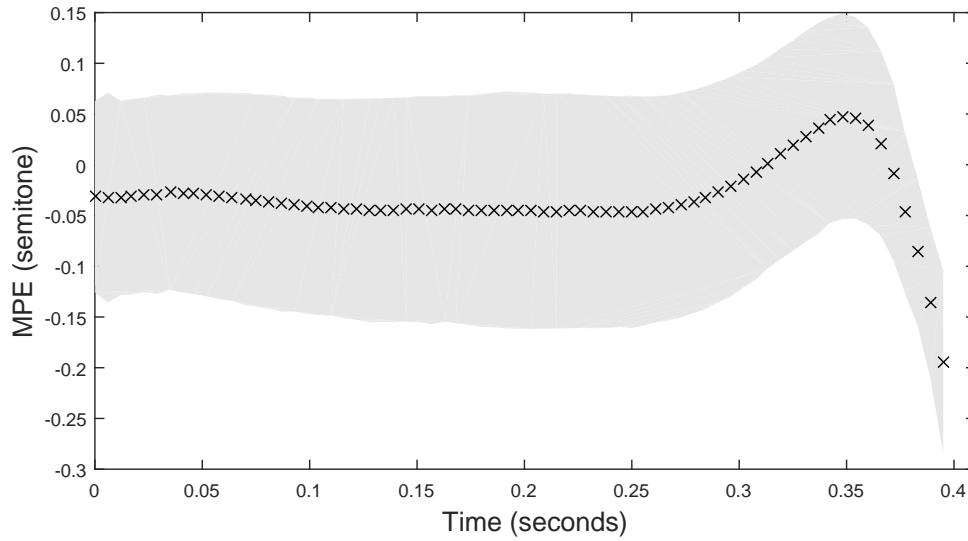
Figure 6.2: Mean pitch error for the final 0.4 seconds of each note where the area shows the standard deviation.

Singers also prepare for the pitch of the next note at the end of each note, as evidenced by the significant difference observed between ascending and descending following intervals ($F(1, 38) = 7.98, p < 0.01$, Figure 6.4). In both cases there is an increase in pitch followed by a rapid decrease as the note ends and the vocal cords are relaxed, but the increase in pitch is much more marked in the case that the succeeding pitch is higher. There are some individual differences between singers in this respect, but most exhibit the average behaviour of being influenced by adjacent notes.
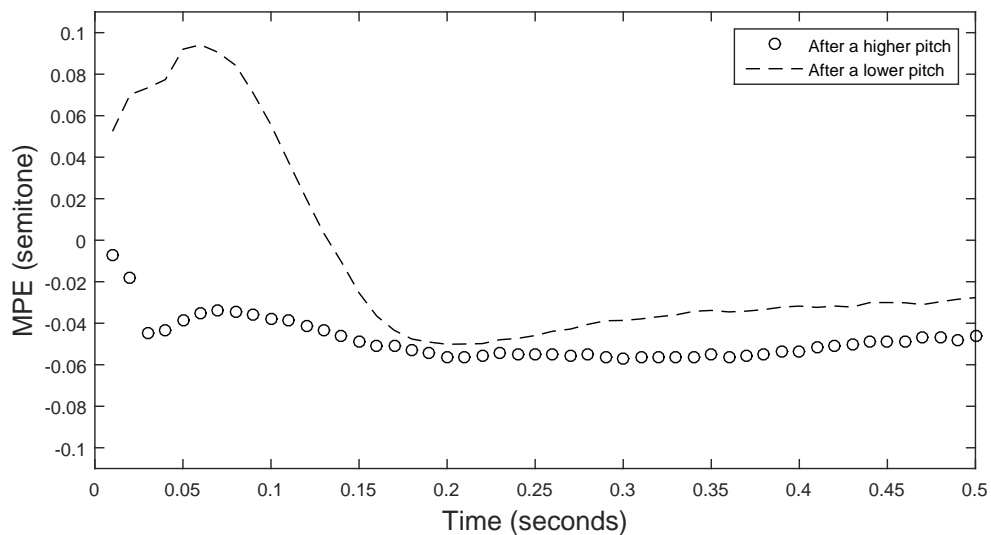
Figure 6.3: The effect of singing after a lower or higher pitch: mean pitch error duration.
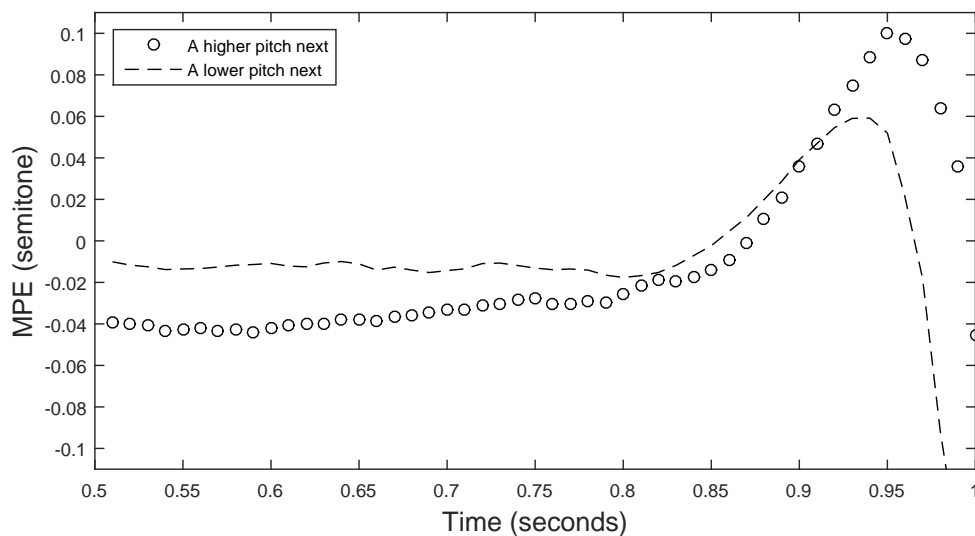
Figure 6.4: The effect of singing before a lower or higher pitch: mean pitch error in real time duration.

6.2.3 Vocal parts and sex

To explore the factor of vocal part, the note trajectories in real time duration of the note initial and final 0.4 seconds were plotted for each of the four vocal parts (Figure 6.5). Firstly, about an 8-cent pitch difference between each pair of adjacent parts was observed in the data. Although the pitch trajectories vary according to the participants, for most participants, sopranos tend to sing sharp while tenors and basses tend to sing flat. These pitch differences lead to an expansion of harmonic intervals between vocal parts, the opposite of the compression that is often observed for melodic intervals (Pfordresher and Brown, 2007).

This phenomenon is also observed between sexes. In general, male participants sing 11 cents flatter while females sing 4 cents sharper than the score pitch. An ANOVA shows a significant difference between the note beginnings of male singers and female singers ($F(1, 198) = 734.99, p < .001$). Male singers tend to begin the note at a higher pitch and adjust downwards, while female singers' initial trajectories have a convex shape, beginning at a lower pitch, overshooting the target, then decreasing toward the target. All the singers tend to have similar note ending, a slight increase in pitch followed by a rapid decrease.

6.2.4 Modelling the note trajectories

For a better understanding of the tendencies of pitch trajectories, the note trajectories were modelled as three separate components: initial transient, note middle and final transient. The tendency of each component was approximated by linear regression. Figure 6.6 shows
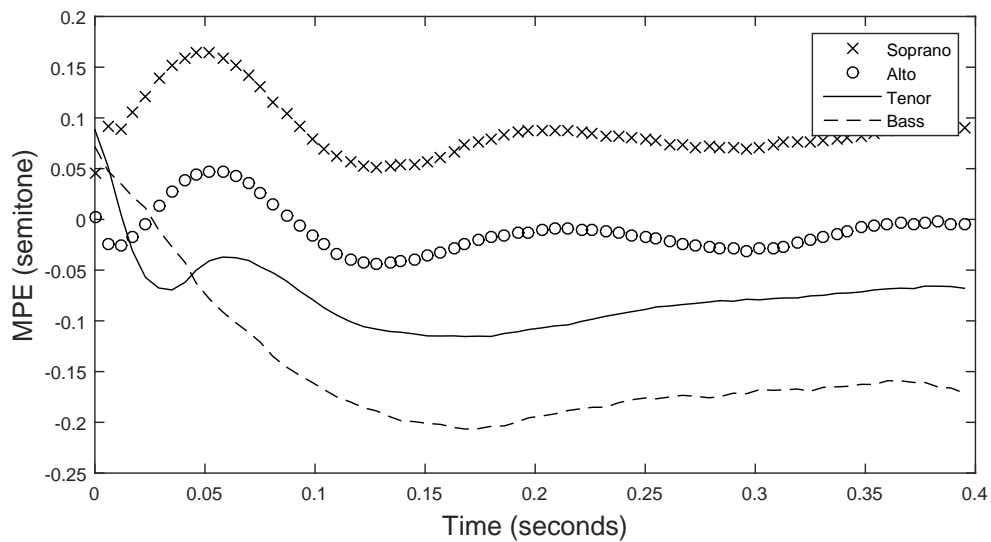
(a) The initial 0.4 seconds
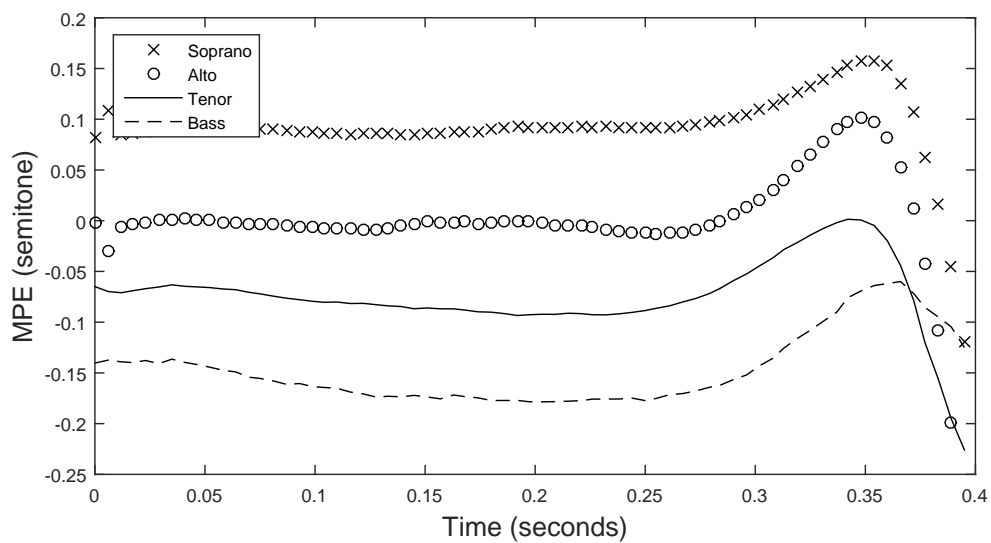
(b) The final 0.4 seconds

Figure 6.5: Mean pitch error for the initial and final 0.4 seconds of each note for each vocal part.

Shape	Attack	Release	Soprano	Alto	Tenor	Bass	Overall
Convex	positive	negative	34.7%	37.8%	22.9%	21.1%	28.9%
Upward	positive	positive	17.1%	13.3%	17.3%	16.1%	16.0%
Downward	negative	negative	32.9%	37.9%	42.4%	33.9%	36.8%
Concave	negative	positive	15.3%	10.9%	17.4%	28.9%	18.4%

Table 6.1: Definition of the four trajectory shapes according to the sign of the slope in the attack and release, and their relative frequencies in each vocal part and in total.

an example of a single pitch trajectory and the linear fits for each of the three components.
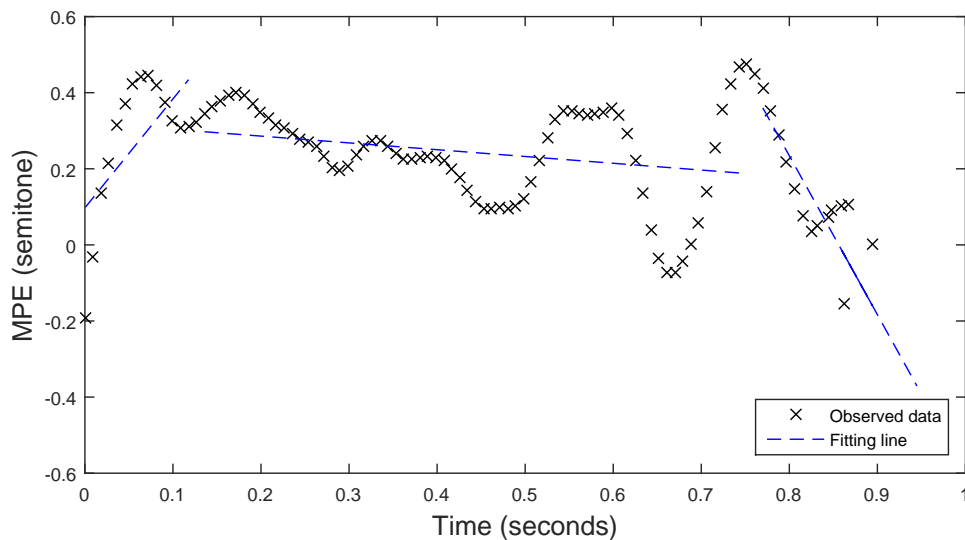
Figure 6.6: Example of the pitch trajectory of a single note and the fitting lines for the initial, middle and final components of the note in real time duration.

To describe the different types of trajectories, they were classified into four categories (Concave, Convex, Upward, Downward) according to the slopes of their initial and final transients, which are either positive or negative as illustrated fin Figure 6.7. Table 6.1 shows that the most frequently occurring shapes are Convex and Downward according to the number of notes of each type, both of which have a negative note release.

Table 6.2 shows the mean, median and standard deviation of the slopes of the three note parts. Although the average trend for the initial transient is a negative slope, less than half of the notes exhibit this behaviour, and there is a large variance in the slope of the initial transient. The middle segment has a small positive trend, while for the final transient most notes have a negative slope, although again this has a large variance.
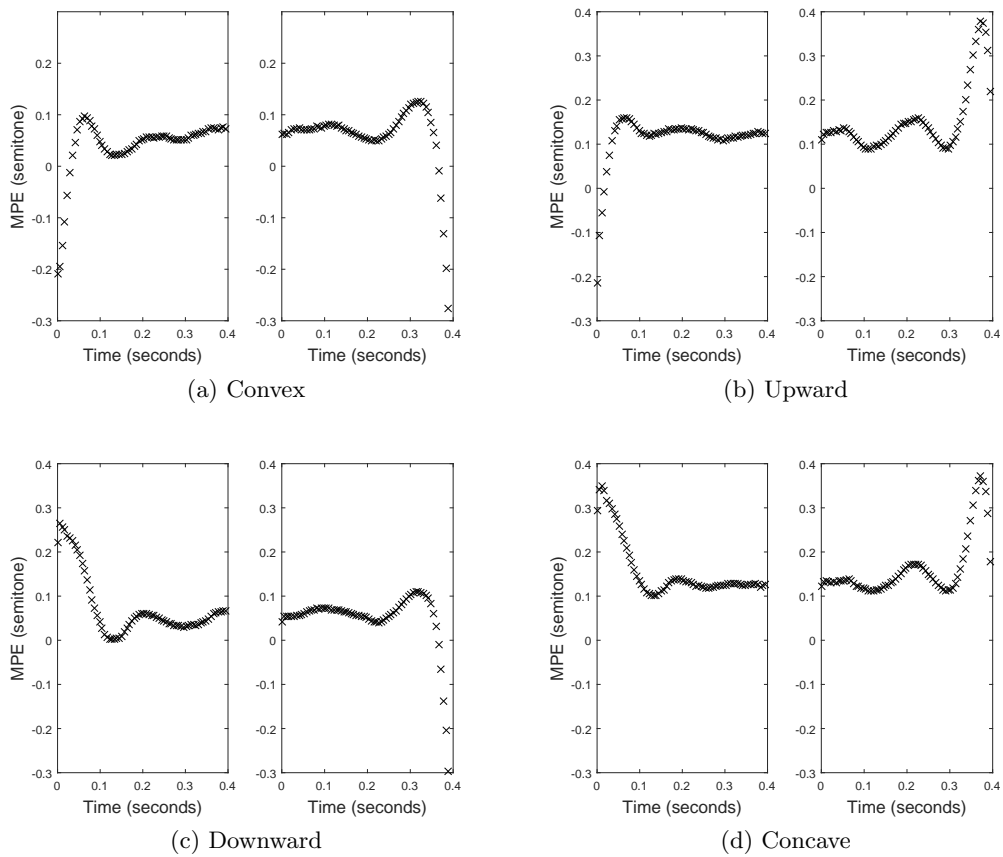
Figure 6.7: Mean pitch trajectories of the four trajectory classes in real time (first 0.4 seconds and last 0.4 seconds of the duration).

	Initial	Middle	Final
Mean	-0.649	0.077	-2.167
Median	0.003	0.038	-1.766
Std.dev	7.109	0.725	6.400

Table 6.2: The mean, median and standard deviation of the slope (semitones per second) of the initial transient, middle section and final transient.

6.2.5   Listening condition

Finally, the influence of listening condition on note trajectory classification was considered. The note trajectories of the independent and dependent conditions in the partial-open condition were separated for the comparison. An ANOVA test on the note trajectories did not show any significant difference between independent and dependent singers in the simplex condition, which means that whether the participants could hear other vocal parts or not did not have a significant effect on the overall shape of their note trajectories.

6.3   Discussion

The results show a general stretching of harmonic intervals between vocal parts, so that the bass part sang flat and the soprano part sharp relative to the other vocal parts. Comparing to the given starting notes, the overall tendency was to sing flat, a tendency which increased over time. Pitch drift has been observed in other experiments (Howard, 2003; Terasawa, 2004; Devaney and Ellis, 2008a), and is typically downward in direction, although upward drift has also been observed.

While the averaged note trajectories, particularly when sorted into categories (Figure 6.7), show quite smooth curves, the individual pitch trajectories exhibit much greater degrees of variation (e.g. Figure 6.6, which is not an extreme example). There is a danger that the features observed in the average curves might be artefacts of the averaging process, and may not occur often, if at all, in the individual instances. For example, Figures 6.1 show a concave shape (a small local minimum) in the first 5% (respectively 0.04 seconds) of the note trajectory. Comparing with Figure 6.5, where the two female vocal parts have different initial trajectories to the two male vocal parts, it is likely that the local minimum arises from averaging the categorically different shapes of the male and female parts. The reason that the end of the note trajectory does not exhibit a similar pattern may be due to the greater frequency of Convex and Downward shapes (28.9% and 36.8% respectively), which both have a negative final slope, across the vocal parts (Table 6.1).

The general tendency of notes ending with a negative slope is observed regardless of whether the next pitch is higher or lower, or which vocal part is considered. Although there is a simple explanation, i.e. the relaxation of the vocal muscles at the end of a note, it is noteworthy that singers show evidence of preparing for a higher following pitch by commencing a rising inflection which is then followed by a falling pitch at the end of the note, which might be thought to negate the preparation. Even in the cases of the Upward

and Concave trajectories, the overall increasing slope toward the end of a note finishes with a few sampling points where the pitch decreases (during the final 3% of the note, Figure 6.7).

A skilled singer is able to coordinate their muscles to achieve synchronised control over multiple vocal parameters. Alongside the pitch changes at the ends of each note, there are also variations in amplitude associated with the start or end of the note, which might make some parts of the transient imperceptible (alternatively, some audible parts may be omitted from analysis due to their low amplitude). The note segmentation (determination of note onset and offset times) is based on the default settings of the software Tony, which segments the pitch track into notes according to changes in pitch and energy (Mauch et al., 2015). Different settings and segmentation strategies may influence the results. The coarse segmentation was checked during annotation. A random sample was checked more closely after results were obtained. This revealed a small fraction of ambiguous cases where the final slope is dominated by vibrato, and thus could be classified as positive or negative, depending on the precise offset time. Compared to the thousands of notes which have a negative slope at the end, the few ambiguous cases would not change the results significantly if they were to be segmented differently.

Although vibrato is a feature of many observed singing pitch trajectories, it is not the main research target in this work. The use of vibrato is less marked in unaccompanied ensemble singing where the voice does not need to be projected over instrumental parts, and the stylistic goal is for the voices to blend rather than stand out. For example, choral style favours minimal vibrato, and barbershop style generally forbids vibrato. Thus strong vibrato was not observed in the data, and in the cases where vibrato was present, it tended to be uneven, which would make it difficult to model.

Since the experiment asked participants to sing /ta/ rather than the lyrics, the consonant at the beginning of the syllable may influence the pitch trajectory within the notes, especially the final transient parts. Hence it is worth to compare the influence of the different syllables. The tenor in the second group sang /da/ rather than /ta/ for all the trials due to the personal preference (audios can be found in Section 6.5). Figure 6.8 shows the note trajectory of the tenor in second group, which is same as the tenor part in Figure 6.5. The bass in the third group also sang syllables /Do/, /Re/, /Mi/ etc. rather than /ta/ in the second piece. The note trajectories are similar to the previous results which tend to end with a downward glide. Although the participant sang different syllables, the transient parts show a similar trajectory. Comparing with the other participants, the syllable

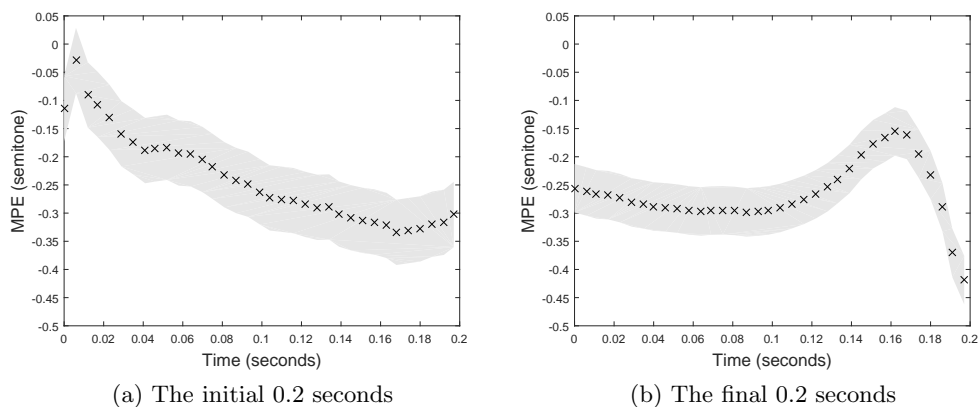(a) The initial 0.2 seconds

(b) The final 0.2 seconds

Figure 6.8: Note trajectory of the tenor in the second group, who sang /da/ rather than /ta/, where the grey area shows the standard deviation.

does not have a significant effect on the pitch trajectory where fundamental frequency was extracted by the software Tony.

6.4 Conclusions

This chapter presents a study of pitch trajectories of single notes in quartet. The main contribution of this chapter is the observation, measurement and analysis of the note transient parts by characterising their shapes and influencing factors. According to the analysis of over 35000 individual notes, a general shape of vocal notes was found which contains transient components at the beginning and end of each note.

The analysis is based on both absolute and relative timing of notes, where the initial and final transients are about 120 ms. The results suggest that the adjustment of pitch at the ends of notes is governed by absolute timing, i.e. due to physiological and psychological factors, rather than relative timing, which might imply a musical motivation. The transient components vary according to the individual performer, previous pitch, next pitch, vocal part and sex.

Participants tend to overshoot the target pitch when transitioning from a lower pitch and raise the pitch toward the end of the note if the next pitch is higher. The results also show a general expansion of harmonic intervals: about 8 cents pitch difference is observed between adjacent vocal parts, with sopranos singing sharper and male singers flatter than the target pitch. Female and male singers also differ in their initial transients, with females commencing with a upward glide that overshoots the target, followed by a correction, while males begin notes with a downward glide. Participants with fine pitch accuracy tend to

have smoother pitch trajectories, while less accurate singers have relatively unstable note trajectories.

6.5 Data availability

The code and the data needed to reproduce the results (note annotations, questionnaire results, score information) are available from `https://code.soundsoftware.ac.uk/projects/analysis-of-interactive-intonation-in-unaccompanied-satb-ensembles/repository`.

Chapter 7

Conclusions and Future Perspectives

7.1 Summary

This thesis discussed three experiments with singers and vocal ensembles to investigate intonation and interaction between singers, with an emphasis on how singers negotiate a joint reference pitch as the music unfolds over time. These three experiments were designed coherently to cover a range of situations where participants sing alone, in duets and quartets; singing conditions are simultaneously and sequenced, unison and duet; and listening conditions, whether singers can hear each other, are solo, partial and open condition.

These experiments address a gap in singing intonation studies, by investigating the effects of interaction between singers on intonation accuracy, as well as other factors, including the score information, the music training background, vocal part, and the presence of a reference pitch (or a live vocal partner). Generally, singing solo or sequenced led to better performance (less pitch error and more stable intonation) than singing with a partner or with a stimulus simultaneously. When singing with a partner, unison singing has better intonation accuracy than singing duet, and smaller melodic intervals and consonant harmonic intervals tended to result in less error. In addition, adjacent pitches and sex of the singer also affect the pitch trajectory and accuracy.

The majority of the work presented in this thesis has been presented at international conferences, as shown in Section 1.4. In addition, all of the data have been published for the use of research community. The main contributions of the thesis are summarised and directions for future work are presented below.

7.1.1 Imitation Study

In order to gain a basic understanding of vocal intonation and interaction, the first experiment uses an imitation task to test how single singers respond to controlled stimuli containing time-varying pitches. 43 participants took part in this experiment, and each imitated 75 instances of five stimulus types in two conditions (simultaneously and sequenced).

According to the linear mixed-effects model, several factors significantly influence the absolute pitch error, which include: stimulus type, main pitch, trial condition, repetition, duration of transient, direction and magnitude of pitch deviation, observed duration, and self-reported musical training and perceptual abilities. The remaining factors that were tested had no significant effect, including self-reported singing ability.

The results show that time-varying stimuli are more difficult to imitate than constant pitches, as measured by absolute pitch error. In a natural performance setting, this suggests that tuning with the human voice is more difficult than with a pitched instrument. In particular, stimuli which end on a pitch other than the main pitch (*tail*, *ramp* and *vibrato* stimuli) had significantly higher absolute pitch errors than the *stable* stimuli, with effect sizes ranging from 4.1 cents (*tail*) to 15 cents (*ramp*).

In addition, listening conditions affect the pitch accuracy by reducing accuracy when other voices are heard. Contrary to expectations, participants performed 3.2 cents worse in the condition when they sang simultaneously with the stimulus, although they also heard the stimulus between singing attempts, as in the sequenced condition. The results are consistent with the duet and quartet experiments.

A positive correlation between the extent of pitch deviation (pitch difference, p_D) and pitch error was found. Although the effect size was small, it was significant and of a similar order to the overall mean pitch error. Although the imitation experiment used a standard main pitch, the later duet and quartet experiments show a significant influence of heard pitch on sung pitch. The results also show that the duration d of the transient proportion of the stimulus correlated with absolute pitch error.

Finally, parameters of the responses were extracted by a forced fit to a model of the stimulus type, to better understand how participants imitated the time-varying stimuli. The resulting curves matched the stimuli more closely than the raw data did.

The imitation experiment shows the influence of several factors where the shape, main pitch, transient part, and listening condition affect the intonation accuracy. The experiment led to a new research question, how singers respond when performing with a live

partner.

7.1.2 Duet Interaction

According to the imitation experiment, various features of the heard intonation affect the sung intonation accuracy, but how does this differ when singing with a vocal partner? Unlike the artificial stimulus which has certain quantitative parameters, the practical circumstance is complex, as it is hard to classify the shape and transient part of human pitch trajectories. Thus the second experiment investigates singing interaction by analysis of the factors influencing pitch accuracy of unaccompanied duet singers.

16 female participants took part in this experiment, where each pair of participants sang two pieces of music in two singing conditions (unison and duet) and three types of listening conditions (solo, simplex, and duplex). The results extend the list of factors of influence found in the imitation experiment, so that singing condition, listening condition, vocal part, note number in trial, score factors and individual factors of the singer all have a significant influence on pitch accuracy.

Like the imitation experiment, where singing simultaneously led to more pitch error than singing sequenced, singing with a partner increased the pitch error and melodic interval error of the individuals, but reduced the harmonic interval error between singers. That is, singers adjust their pitch to harmonise better with their partner, but sacrifice their individual accuracy.

In terms of singing conditions, the unison condition has better intonation accuracy than the duet condition. This gives some measure of the additional difficulty of singing in harmony, and particularly of tuning non-unison intervals.

The target harmonic interval has a significant effect on harmonic interval error, with dissonant intervals having the largest errors and the unison interval the smallest. In the preliminary project, the tritone (6 semitones) interval was the most difficult melodic interval. As for harmonic intervals, the least consonant intervals have the greatest error, with the minor second and tritone having the largest harmonic interval error and also the largest spread of values.

A positive correlation between the signed pitch errors of dependent singers and independent singers in the simplex condition was found. In other words, if one singer sings sharp, their partner is influenced to sing sharp as well. Similar results were found in both the imitation and quartet experiments. The correlation of pitch errors is again evidence of interaction, that singers adjust their pitch to improve harmonic intervals at the expense

of melodic intervals and of preservation of the tonal reference.

Analysis of the pitch trajectories within tones revealed greater stability of pitch in the unison condition than the duet condition, but not in independent singers over dependent singers. The results suggest that the observed pitch variation arises more from imprecision or uncertainty than deliberate adjustment.

7.1.3 SATB Interaction

The third experiment extends the study of singing conditions and listening conditions to the case of interaction in four-part (SATB) singing. An experiment with 20 participants (five groups of four) was conducted. Each group sang two pieces of music in three listening conditions: solo, with one vocal part missing and with all vocal parts.

The results confirm that interaction exists between singers and influences their intonation once again; moreover, intonation accuracy depends on which other singers each individual singer can hear. In particular, singers were more accurate when they could not hear the bass. This surprising result might be due to the fact that the bass singers were less accurate on average than other singers in this experiment. In the subsequent intonation trajectory study, the bass part also tended to sing flat.

The results revealed increases in pitch error and melodic interval error were observed when singers could hear each other. Singers have the best intonation accuracy when they sing alone without any vocal accompaniment from other vocal parts. The more vocal accompaniment they have (with one or more other parts), the less intonation accuracy they have. At the same time, hearing their partner decreased the harmonic interval error, while the direction of the communication (one-way or two-way) did not influence the harmonic interval error.

Interaction also has the effect on the note stability of increasing its variability. Pitch within a note varies more when singers hear each other; as one might expect the singers are adjusting their intonation to be in tune with each other.

7.1.4 Intonation Trajectories

Besides the interaction and intonation of complete notes, the pitch variation inside each single note is another interesting topic. The main contribution of this final study is the observation, measurement and analysis of the note transient parts, characterising their shapes and influencing factors.

Over 35000 pitch trajectories of single notes from the SATB experiment were analysed

and a general shape of vocal notes was found which contains transient components at the beginning and end of each note.

According to the analysis of within-note pitch variation, the initial and final transients are about 120 ms, which is similar to the previous modelling results and experiment design of the imitation experiment. Although the initial transients may have different directions, the final transients tend to decrease in pitch. This may due to physiological and psychological reasons. The transient components vary according to the individual performer, previous pitch, next pitch, vocal part, and sex. More specifically, when the target pitch is higher than current pitch, participants tend to overshoot the pitch with a pitch rise in the initial transients.

Different vocal parts show tendencies to sing sharp or flat over all; where sopranos sing sharper, tenor and bass sing respectively 8 and 16 cents flatter than the soprano pitch. Female and male singers also differ in their initial transients, with females commencing with an upward glide that overshoots the target, followed by a correction, while males begin notes with a downward glide. Participants with fine pitch accuracy tend to have smoother pitch trajectories, while less accurate singers have relatively unstable note trajectories.

7.2   Future Perspectives

In the process of working on the experiments for this thesis, and the writing of the thesis itself, many exciting new research ideas have arisen, however, due to the time limitation and practical reasons, I'd like to implement these idea after this thesis. Here, I would like to mention a selection of ideas for future work on intonation and interaction.

7.2.1   Multiple singers for each vocal part

In the experiments of this thesis, the reaction of a single singer to a stimulus, and the interaction of duet and quartet ensembles have been investigated. However, choirs usually have multiple singers for each vocal part. In order to move towards more typical musical settings, further studies need to investigate cases where there are multiple (more than two) singers per part. To extend the study of SATB singing in Chapter 5, an experiment with multiple singers in each of the four vocal parts could be concluded.

The results in this thesis have shown that interaction significantly influences pitch accuracy, singing with multiple vocal parts leading to increases in the pitch error, melodic interval error, note stability and even leading to pitch drift. Further research could focus on how to improve the pitch accuracy and prevent pitch drift as observed in these exper-
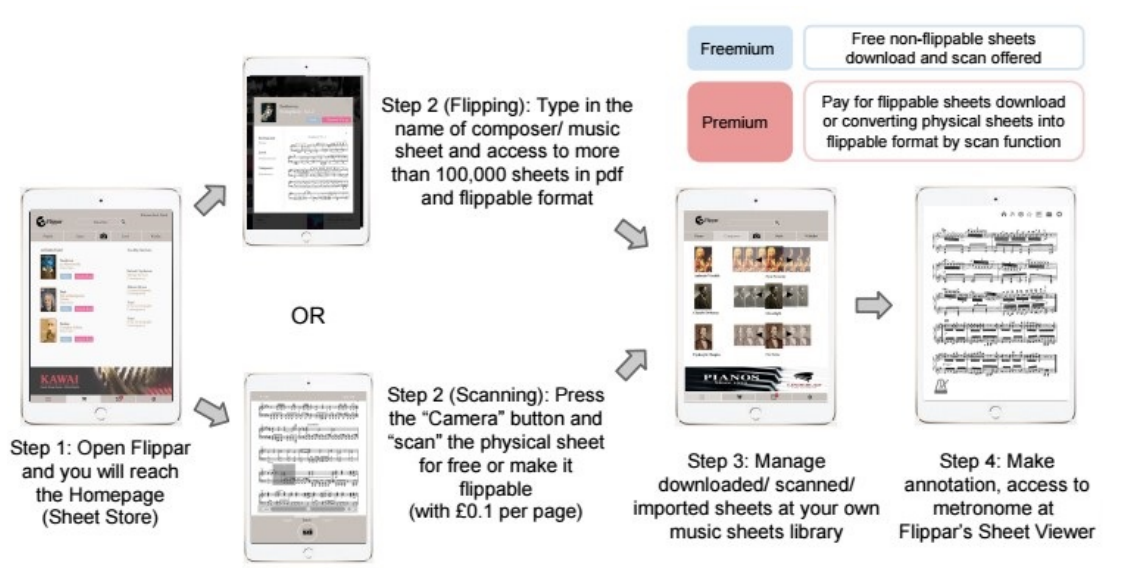
Figure 7.1: The interface design of Flippar

iments, since the main target of this work has been testing and comparing the intonation accuracy without consideration of the special training processes.

For example, researchers could recruit two groups of participants (12 singers each group and 3 singers each vocal part) and regularly record the performance of the participants each week, where one group sing without interventions and another group with different listening condition and singing condition.

Moreover, here are some rehearsal suggestions as found in singing pedagogy literature: 1) The singers having relatively worse pitch accuracy require a higher-density training. 2) Singers need to learn and practice the piece individually and train to hear themselves. 3) All the singers are trained to shorten their transients parts when reaching a target pitch. 4) The rehearsal starts with separate whole ensembles into two groups according to sex before the ensemble. 5) Aligning with some vocal part has more priority than sing harmonic with other vocal parts.

7.2.2 Flippar: a real time intonation accuracy App

Flippar is the name of a proposed new application that allows musicians to play music without having to manually turn pages of music. The author's business proposal for Flippar won the first prize of the 7th innovation and entrepreneurship competition in the UK area (July 2015), however, the proposal has not been realised yet. Figure 7.1 shows the interface and instructions for Flippar.

There are some efficient existing software packages for annotating pitch and note information; however, users need to import the audio files after recording and process them

offline. A smart phone application which shows the intonation accuracy simultaneously would be a great assistant for vocal ensembles and instrument players.

The idea behind Flippar is to combine both pitch extraction and page-turning functions. Using the pYIN algorithm, the application is able to recognise the fundamental frequency of the note and flip over the digital music sheets automatically instead of the musicians rushing to turn over the sheets by themselves. Users need to pay for the copyright of protected music, but they are free to use their own uploaded score if not subject to copyright.

Although some karaoke applications implement the function of pitch extraction, there is only one direct competitor which offers the automatic music sheet turning solution in the global context, and it uses another technology, called beat tracking, instead of pitch recognition to follow the music notes.

Using the Flippar application, users can scan sheet music and convert it to MIDI format using optical music recognition (OMR) technology. When the users sing or play the music, the application automatically turns the pages and shows the pitch accuracy. The scanned music will be added to a database for sharing and cooperation. Especially in the vocal module, the singers can observe their intonation accuracy simultaneously and adjust their intonation. The module will also give a personal intonation analysis for each user according to their user record.

7.3 Study of singing techniques

The experiments in this thesis focus on the intonation accuracy and interaction of vocal ensembles. However, a pitch-accurate singer is not equivalent to a good singer. Although intonation accuracy is one of the key principles, in practical performance, the tone and singing technique are also crucial. Some singers who have a great singing technique may be measured as inaccurate in many cases. For example, a soprano in the opera produces a wide vibrato, which may lead to high pitch variance and an apparently incorrect main pitch.

In future work, I would like to explore the relationship between singing technique and the evaluation of vocal quality. Although singing technique is not the focus of this thesis, it is of great interest to investigate how singers produce a certain sound with their tonal memory and muscle movement, and how the listeners evaluate the vocal quality. The idea was inspired by an interesting personal observation: some famous Asian singers are evaluated as "out of tune" by Western listeners when they sing folk music with special

singing techniques. Listeners from a different cultural background may think differently. Thus any evaluation of singing quality would need to take into account not just pitch, but also timbre, dynamics, articulation, and the cultural context of the singing.

Appendix A

Online System of Data Collection for Imitation Study

A.1 Online Data Collection

The basic function of this module is to collect the test audio according to the participants' sex, keep a record of test information and upload the data to the server. Only registered singers can log in to this panel.

There are in total 2(test) × 75(stimulus) × 2(sex)=300 audio files in the stimulus pool. The pieces were played according to the singers' sex and randomly shuffle within one test. The system generates a play order for 75 notes and saves this order in the user database for the analysis. Whenever the singer starts the experiment, the system reads the index of the unfinished tests and plays these test pieces in the specific order.

A.1.1 Playing Recording Function

The experiment panel has two sections as shown in Figure A.1. On the left are control buttons and basic information while the playing interface is on the right. When participants are ready to record, press the "Recording" button, and the system will choose the required audio and video, and play them simultaneously. The bottom right shows the experiment status (recording, encoding, uploading and whether the test is finished). If participants are not satisfied with the most recent recording, they can repeat the recording by pressing the "Rec Again" button, otherwise they go to the next note by pressing the "Next" button. Participants can also watch the tutorial video and fill in the questionnaire with this panel. The recording and uploading functions are completed in the background. If any problem happens, participants only need to refresh the page and no data loss will
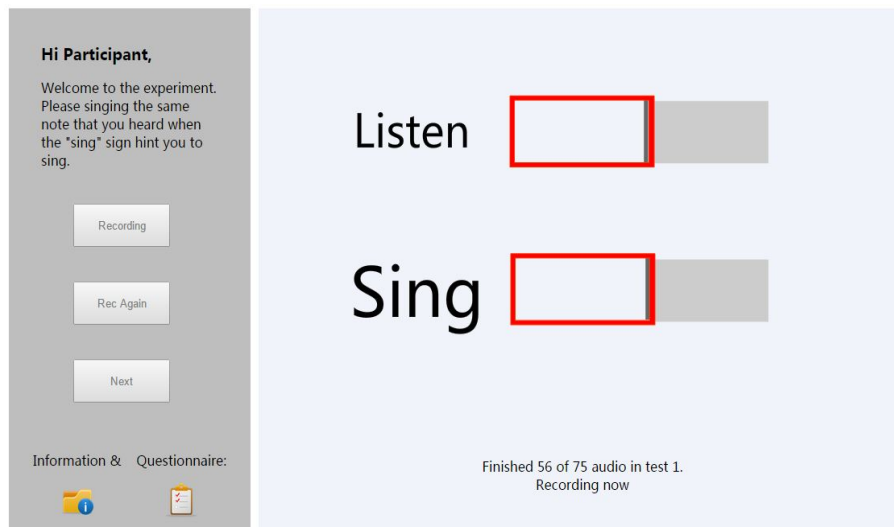
Figure A.1: The interface of online system.

occur.

The audio and video are played when they are loaded. Because there are 300 audio files and 2 video files, the page preloads the media materials to optimise the loading time. After loading, the system will play the media automatically. There are two red rectangles in the interface; the upper one indicates when the audio is playing and the lower one indicates when the participants need to sing. A monochrome rectangle will align with the red rectangle exactly when audio files are played or singers sing. The participants can take a break between trials whenever they want. If the recorded audio file was not uploaded or the singer is interrupted, they can record the unfinished trial again. "Information" and "Questionnaire" options link to the tutorial video and online questionnaire.

There were some difficulties in the website design. In the earliest prototype, if participants pressed the keyboard rapidly, it sometimes blocked the browser. Therefore the Next button is disabled until the audio has been played. For security reasons, the browser will ask permission to use the microphone whenever the page is loaded. To avoid this pop-up occurring for every trial, the researcher used jQuery AJAX to reload just the right side of the page. The AJAX function will call a PHP script in the server to save the index and upload recorded files. The return value of the trial number and test number is printed in the bottom right as the trial is finished.

The recording function was achieved by the getUserMedia interface and Web Audio which allowed to the microphone input to be recorded and exported to a .wav file. A single channel was used to keep the audio file small.

A.1.2   Auto-name & Auto Upload

The recording is named according to the singer's number, test number, test order. For example, if participant P01 is performing test 1, and the system plays note number 30 which is the 56th trial in the set random order, the recorded file will be named as "P01-1-30-56.wav". These file names contain the singer number, test number, note number which are used for the next step: data annotation.

Every singer has 152 separate recordings (150 trials + 2 test sample). Because of some initial bugs and network connection issues, the researcher lost 36 recordings among the first 300 files. After gathering more experience and making the system more robust, the issue was solved. For each singer, the data size is above 300 MB.

A.1.3   Tutorial Video

All the singers were asked to watch a tutorial video to guide them how to use the experiment system. The purpose of this design is to make sure every participant gets the same introduction. The tutorial contains a screen capture of the interface to illustrate its functions. The tutorial is on `https://www.youtube.com/watch?v=xadECsaglHk`.

A.2   On-line Management System

Besides the system for participants, there is another system for the researcher to manage the participants' information and their recordings. This system has a relatively simple interface but is very important for data collecting and analysis.

A.2.1   User Management

The user management system has four basic functions: (1) insert new participants; (2) search the basic information of a particular singer; (3) search singers that finished the experiment; (4) change the experiment index for the particular singer. Figure A.2 shows the interface of the management system.

Because all the participants were pre-selected according to their musical background and their age, not all the users who sign up online can be accepted. The selection depends on the availability and competency of the participants.

Only invited participants can join the experiment.

When the researcher retrieves the information of a certain singer, the system will output MATLAB code for that singer which includes name, sex, test order and email address (the

Figure A.2: Interface of Management System.

data was analysed anonymous, the name and personal information is for the researcher to retrieve feedback from participants). The researcher can use the code directly for further analysis in MATLAB.
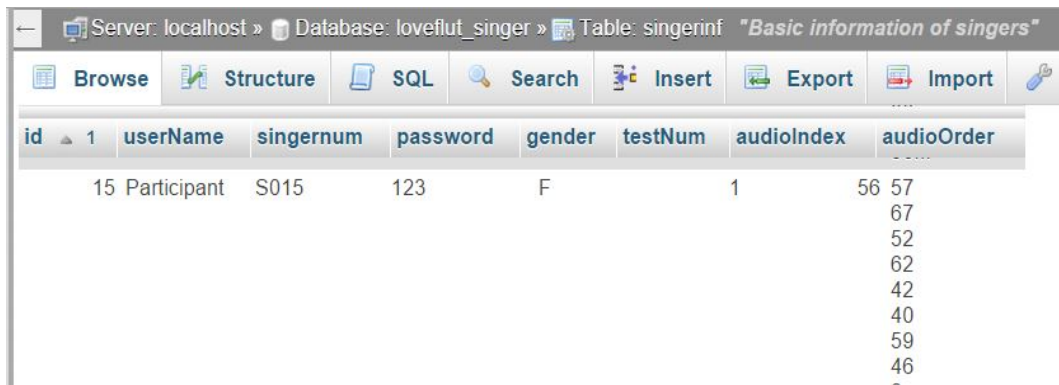
The "Search finished singer" function returns basic statistical results of finished singers. It is useful when recruiting the next participants. For example, if from the first 10 participants the search results showed 8 of them are female, the author should find more male singers as the next participants.

The last function is for singers who make a mistake in the experiment. For example, if a singer presses the "Next" button without singing anything, or she records something else rather than the notes that the researcher asked her to sing, the researcher will be notified by the system and the participants will be asked to finish the missing trial. Another advantage of the function is that it saves writing SQL commands and allows researchers to modify the data more easily. To be more specific, the management function can be accessed from any device, such as a mobile phone, thus the researcher could monitor the participants remotely.

A.3   Data Storage

All the raw data are stored in the SQL system and server. Figure A.3 shows the raw information of the singers. The participants can take breaks and log in to the experiment whenever they are available after registering. The personal data are saved as participants are selected. The audio play order is generated automatically and then does not change any further. The test index and audio play index increase when data are uploaded from

the browser until the experiment is finished.

Figure A.3: The SQL panel.

Taking the singer in Figure A.3 as an example, once the singer logs in to the system and inputs her password, the system will check the username and password, and since she selected her sex as "F", the system will play the experiment audio files for the female in the order of file number 57, 67, 52, etc. Every time she finishes a trial, the audio index will increase until she finishes the experiment. At the end of the experiment, the system checks whether the singer finished the questionnaire or not.

All the data for every singer were saved in a structure as shown in Figure A.4. In this structure, all the note arrays can be searched and retrieved. All the structures were saved in MATLAB .mat files for further analysis.

singer
├── .name - 'name'
├── .gender - 'M/F'
├── .order - 75*1 array
├── .singerNum - string
├── .test1 ┬ .note1 ┬ .first
│   ├ .note2 ├ .second
│   ├ ...   └ .third
│   └ .note75
└── .test2 ┬ .note1 ┬ .first
           ├ .note2 ├ .second
           ├ ...   └ .third
           └ .note75
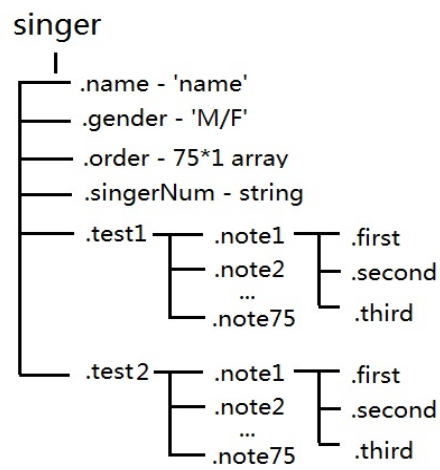
Figure A.4: Data Structure.

Appendix B

Music pieces of the Experiments

B.1   Music pieces of the preliminary project

Do-Re-Mi

Music by Richard Rodgers

Edelweiss

Music by Richard Rodgers

My Favorite Thing

Music by Richard Rodgers

B.2 Music pieces of the SATB experiment

Oh Thou, Of God The Father

Bach

To be or not to be

Evling Zullberg

Leo Mathisen

Appendix C

Questionnaire

In the experiments of Chapter 2, 3, 4 and 5 participants completed the questionnaire containing 34 questions from the Goldsmiths Musical Sophistication Index (Müllensiefen et al., 2014) which can be grouped into 4 main factors for analysis. The following questions are the questions answered by the participants, where the number before the question is the order in the questionnaire and * is optional questions.

Singer Number *

Please input your singer number

Your gender*

A Female

B Male

Your Nationality*:

Country of Current Residency:

Your age *

Email address*

Factor 1 - Active Engagement

1. I spend a lot of my free time doing music-related activities.

A Completely Disagree

B Strongly Disagree

C Disagree

D Neither Agree nor Disagree

E Agree F Strongly Agree

G Completely Agree

2. I enjoy writing about music, for example on blogs and forums.

A Completely Disagree

B Strongly Disagree

C Disagree

D Neither Agree nor Disagree

E Agree F Strongly Agree

G Completely Agree

7. I'm intrigued by musical styles I'm not familiar with and want to find out more.

A Completely Disagree

B Strongly Disagree

C Disagree

D Neither Agree nor Disagree

E Agree

F Strongly Agree

G Completely Agree

13. I often read or search the internet for things related to music.

A Completely Disagree

B Strongly Disagree

C Disagree

D Neither Agree nor Disagree

E Agree

F Strongly Agree

G Completely Agree

16. I don't spend much of my disposable income on music.

A Completely Disagree

B Strongly Disagree

C Disagree

D Neither Agree nor Disagree

E Agree

F Strongly Agree

G Completely Agree

19. Music is kind of an addiction for me - I couldn't live without it.

A Completely Disagree

B Strongly Disagree

C Disagree

D Neither Agree nor Disagree

E Agree

F Strongly Agree

G Completely Agree

23. I keep track of new music that I come across (e.g. new artists or recordings).

A Completely Disagree

B Strongly Disagree

C Disagree

D Neither Agree nor Disagree

E Agree

F Strongly Agree

G Completely Agree

28. I have attended ( ) live music events as an audience member in the past twelve months.

A 0

B 1

C 2

D 3

E 4-5

F 6-9

G 10 or more

32. I listen attentively to music for ( ) per day.

A 0-15 minutes

B 15-30 minutes

C 30-60 minutes

D 60-90 minutes

E 2 hours

F 2-3 hours

G 4 hours or more

Factor 2 - Perceptual Abilities

4. I am able to judge whether someone is a good singer or not.

A Completely Disagree

B Strongly Disagree

C Disagree

D Neither Agree nor Disagree

E Agree

F Strongly Agree

G Completely Agree

5. I usually know when I'm hearing a song for the first time.

A Completely Disagree

B Strongly Disagree

C Disagree

D Neither Agree nor Disagree

E Agree

F Strongly Agree

G Completely Agree

9. I find it difficult to spot mistakes in a performance of a song even if I know the tune.

A Completely Disagree

B Strongly Disagree

C Disagree

D Neither Agree nor Disagree

E Agree

F Strongly Agree

G Completely Agree

10. I can compare and discuss differences between two performances or versions of the same piece of music.

A Completely Disagree

B Strongly Disagree

C Disagree

D Neither Agree nor Disagree

E Agree

F Strongly Agree

G Completely Agree

11. I have trouble recognising a familiar song when played in a different way or by a different performer.

A Completely Disagree

B Strongly Disagree

C Disagree

D Neither Agree nor Disagree

E Agree

F Strongly Agree

G Completely Agree

15. I can tell when people sing or play out of time with the beat.

A Completely Disagree

B Strongly Disagree

C Disagree

D Neither Agree nor Disagree

E Agree

F Strongly Agree

G Completely Agree

17. I can tell when people sing or play out of tune.

A Completely Disagree

B Strongly Disagree

C Disagree

D Neither Agree nor Disagree

E Agree

F Strongly Agree

G Completely Agree

18. When I sing, I have no idea whether I'm in tune or not.

A Completely Disagree

B Strongly Disagree

C Disagree

D Neither Agree nor Disagree

E Agree

F Strongly Agree

G Completely Agree

21. When I hear a piece of music I can usually identify its genre.

A Completely Disagree

B Strongly Disagree

C Disagree

D Neither Agree nor Disagree

E Agree

F Strongly Agree

G Completely Agree

Factor 3 - Musical Training

12. I have never been complimented for my talents as a musical performer.

A Completely Disagree

B Strongly Disagree

C Disagree

D Neither Agree nor Disagree

E Agree

F Strongly Agree

G Completely Agree

22. I would not consider myself a musician.

A Completely Disagree

B Strongly Disagree

C Disagree

D Neither Agree nor Disagree

E Agree

F Strongly Agree

G Completely Agree

26. I engaged in regular, daily practice of a musical instrument (including voice) for ( ) years.

A 0

B 1

C 2

D 3

E 4-5

F 6-9

G 10 or more

27. At the peak of my interest, I practised ( ) hours per day on my primary instrument.

A 0

B 0.5

C 1.0

D 1.5

E 2.0

F 3-4

G 5 or more

29. I have had formal training in music theory for ( ) years.

A 0

B 0.5

C 1

D 2

E 3

F 4-6

G 7 or more

30. I have had ( ) years of formal training on a musical instrument (including voice) during my lifetime. *

A 0

B 0.5

C 1

D 2

E 3-5

F 6-9

G 10 or more

31. I can play ( ) musical instruments.

A 0

B 1

C 2

D 3

E 4

F 5

G 6 or more

Factor 4 - Singing Abilities

3. If somebody starts singing a song I don't know, I can usually join in.

A Completely Disagree

B Strongly Disagree

C Disagree

D Neither Agree nor Disagree

E Agree

F Strongly Agree

G Completely Agree

6. I can sing or play music from memory.

A Completely Disagree

B Strongly Disagree

C Disagree

D Neither Agree nor Disagree

E Agree

F Strongly Agree

G Completely Agree

8. I am able to hit the right notes when I sing along with a recording.

A Completely Disagree

B Strongly Disagree

C Disagree

D Neither Agree nor Disagree

E Agree

F Strongly Agree

G Completely Agree

14. I am not able to sing in harmony when somebody is singing a familiar tune.

A Completely Disagree

B Strongly Disagree

C Disagree

D Neither Agree nor Disagree

E Agree

F Strongly Agree

G Completely Agree

20. I don't like singing in public because I'm afraid that I would sing wrong notes.

A Completely Disagree

B Strongly Disagree

C Disagree

D Neither Agree nor Disagree

E Agree

F Strongly Agree

G Completely Agree

24. After hearing a new song two or three times, I can usually sing it by myself.

A Completely Disagree

B Strongly Disagree

C Disagree

D Neither Agree nor Disagree

E Agree

F Strongly Agree

G Completely Agree

25. I only need to hear a new tune once and I can sing it back hours later.

A Completely Disagree

B Strongly Disagree

C Disagree

D Neither Agree nor Disagree

E Agree

F Strongly Agree

G Completely Agree

I have sung in choirs for —- years. *

A 0

B 0.5

C 1

D 2

E 3-5

F 6-9

G 10 or more

I have absolute pitch*.

A No.

B Don't know.

C Yes.

Bibliography

Alldahl, P. (2006). *Choral Intonation.* Gehrman, Stockholm, Sweden. p. 4.

Alldahl, P.-G. (2008). *Choral Intonation.* Gehrmans.

American National Standards Institute (1973). *American National Standard Psychoacoustical Terminology. S3.20.* American National Standards Institute, New York.

Baharloo, S., Service, S. K., Risch, N., Gitschier, J., and Freimer, N. B. (2000). Familial aggregation of absolute pitch. *The American Journal of Human Genetics*, 67(3):755–758.

Baken, R. J. and Orlikoff, R. F. (2000). *Clinical Measurement of Speech and Voice.* Cengage Learning.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48.

Benetos, E. (2012). *Automatic transcription of polyphonic music exploiting temporal evolution.* PhD thesis, Queen Mary University of London, Centre for Digital Music.

Benward, B. (2014). *Music in Theory and Practice Volume 1.* McGraw-Hill Higher Education.

Berkowska, M. and Dalla Bella, S. (2009). Acquired and congenital disorders of sung performance: A review. *Adv. Cogn. Psychol.*, 5(-1):69–83.

Bohrer, J. C. S. (2002). Intonational Strategies in Ensemble Singing. *Doctoral thesis, City University London.*

Brandler, B. J. and Peynircioglu, Z. F. (2015). A Comparison of the Efficacy of Individual and Collaborative Music Learning in Ensemble Rehearsals. *Journal of Research in Music Education*, 63(3):281–297.

Brown, D. E. (1991). *Human universals.* McGraw-Hill New York.

Brown, W. A., Cammuso, K., Sachs, H., Winklosky, B., Mullane, J., Bernier, R., Svenson, S., Arin, D., Rosen-Sheidley, B., and Folstein, S. E. (2003). Autism-related language, personality, and cognition in people with absolute pitch: results of a preliminary study. *Journal of Autism and Developmental Disorders*, 33(2):163–167.

Burnett, T. A., Freedland, M. B., Larson, C. R., and Hain, T. C. (1998). Voice F0 responses to manipulations in pitch feedback. *Journal of the Acoustical Society of America*, 103(6):3153–3161.

Clayton, L. (1986). An investigation of the effect of a simultaneous pitch stimulus on vocal pitch accuracy. Master's thesis, Indiana University, Bloomington.

Cooper, N. A. (1995). Children's singing accuracy as a function of grade level, gender, and individual versus unison singing. *J. Res. Music Educ.*, 43(3):222–231.

Crowther, D. S. (2003). *Key Choral Concepts: Teaching Techniques and Tools to Help Your Choir Sound Great!* Horizon Publishers, Springville, Utah. pp. 81–85.

Dai, J. and Dixon, S. (2016). Analysis of Vocal Imitations of Pitch Trajectories. In *17th International Society for Music Information Retrieval Conference*, pages 87–93.

Dai, J. and Dixon, S. (2017). Analysis of Interactive Intonation in Unaccompanied SATB Ensembles. In *18th International Society for Music Information Retrieval Conference*, pages 599–605.

Dai, J. and Dixon, S. (2019a). Intonation Trajectories in Unaccompanied SATB Singing. .

Dai, J. and Dixon, S. (2019b). Singing Together: Pitch Accuracy and Interaction in Unaccompanied Duet Singing. *The Journal of the Acoustical Society of America*, 145(2):663–675.

Dai, J. and Dixon, S. (2019c). Understanding Intonation Trajectories and Patterns of Vocal Notes. In *25th International Conference on MultiMedia Modeling*, pages 243–253.

Dai, J., Mauch, M., and Dixon, S. (2015). Analysis of Intonation Trajectories in Solo Singing. In *16th International Society for Music Information Retrieval Conference*, pages 420–426.

Dalla Bella, S., Giguère, J.-F., and Peretz, I. (2007). Singing proficiency in the general population. *J. Acoust. Soc. Am.*, 121(2):1182–1189.

de Cheveigné, A. and Kawahara, H. (2002). YIN, a fundamental frequency estimator for speech and music. *Journal of the Acoustical Society of America*, 111(4):1917–1930.

Demorest, S. M., Pfordresher, P. Q., Dalla Bella, S., Hutchins, S., Loui, P., Rutkowski, J., and Welch, G. F. (2015). Methodological perspectives on singing accuracy: An introduction to the special issue on singing accuracy (part 2). *Music Percept.*, 32(3):266–271.

Devaney, J. and Ellis, D. P. (2008a). An Empirical Approach to Studying Intonation Tendencies in Polyphonic Vocal Performances. *J. Interdiscipl. Music Stud.*, 2(1&2):141–156.

Devaney, J. and Ellis, D. P. W. (2008b). An empirical approach to studying intonation tendencies in polyphonic vocal performances. *J. Interdiscipl. Music Stud.*, 2(1):141–156.

Dromey, C., Carter, N., and Hopkin, A. (2003). Vibrato rate adjustment. *Journal of Voice*, 17(2):168–178.

Fischer, P.-M. (1993). Die stimme des sängers. *Analyse ihrer Funktion und Leistung-Geschichte und Methodik der Stimmbildung, Stuttgart.*

Fyk, J. (1985). Pitch-matching ability in children as a function of sound duration. *Bulletin of the Council for Research in Music Education*, pages 76–89.

Fyk, J. (1995). *Melodic Intonation, Psychoacoustics and the Violin*. Organon.

Ganschow, C. M. (2013). Secondary school choral conductors' self-reported beliefs and behaviors related to fundamental choral elements and rehearsal approaches. *Journal of Music Teacher Education*, 20(10):1–10.

Gerhard, D. (2005). Pitch Track Target Deviation in Natural Singing. In *ISMIR*, pages 514–519.

Goetze, M. (1985). *Factors Affecting Accuracy in Children's Singing*. University of Colorado at Boulder.

Goetze, M. (1989). A comparison of the pitch accuracy of group and individual singing in young children. *Bull. Counc. Res. Music Educ.*, pages 57–73.

Green, G. A. (1994). Unison versus individual singing and elementary students' vocal pitch accuracy. *J. Res. Music Educ.*, 42(2):105–114.

Heath, C. and Gonzalez, R. (1995). Interaction with others increases decision confidence but not decision quality: Evidence against information collection views of interactive decision making. *Organ. Behav. Hum. Decis. Process.*, 61(3):305–326.

Hewitt, M. (2008). *Music Theory for Computer Musicians.* Nelson Education.

Howard, D. (2007a). Equal or non-equal temperament in *A Capella* SATB singing. *Logopedics Phoniatrics Vocology*, 32(2):87–94.

Howard, D. M. (2003). *A Capella* SATB quartet in-tune singing: Evidence of intonation shift. In *Stockholm Music Acoust. Conf.*, volume 2, pages 462–466.

Howard, D. M. (2007b). Intonation drift in *A Capella* soprano, alto, tenor, bass quartet singing with key modulation. *J. Voice*, 21(3):300–315.

Howard, D. M. (2007c). Intonation drift in *A Capella* soprano, alto, tenor, bass quartet singing with key modulation. *J. Voice*, 21(3):300–315.

Howard, D. M. and Angus, J. (2017). *Acoustics and Psychoacoustics.* Focal press.

Kalin, G. (2005). Formant frequency adjustment in barbershop quartet singing. *Stockholm: KTH Royal Institute of Technology.*

Kennedy, M. (1980). *The Concise Oxford Dictionary of Music.* Oxford University Press, Oxford, United Kingdom. p. 319.

King, J. B. and Horii, Y. (1993). Vocal matching of frequency modulation in synthesized vowels. *Journal of Voice*, 7(2):151–159.

La Barbara, J. (2002). Voice is the Original Instrument. *Contemporary Music Review*, 21(1):35–48.

Liimola, H. (2000). Some notes on choral singing. In Potter, J., editor, *The Cambridge Companion to Singing*, pages 149–157. Cambridge University Press, Cambridge, UK.

Lindley, M. (2001). Just intonation. *Grove Music Online, edited by L. Macy. http://www.grovemusic.com (accessed 30 January 2015).*

Loeffler, D. B. (2006). *Instrument Timbres and Pitch Estimation in Polyphonic Music.* PhD thesis, Georgia Institute of Technology.

Long, P. A. (1977). Relationships between pitch memory in short melodies and selected factors. *J. Res. Music Educ.*, 25(4):272–282.

Mauch, M. (2010). *Automatic Chord Transcription from Audio Using Computational Models of Musical Context*. PhD thesis, Queen Mary University of London, Centre for Digital Music.

Mauch, M., Cannam, C., Bittner, R., Fazekas, G., Salamon, J., Bello, J., Dai, J., and Dixon, S. (2015). Tony: a Tool for Efficient Computer-aided Melody Note Transcription. In *the First International Conference on Technologies for Music Notation and Representation (TENOR)*.

Mauch, M. and Dixon, S. (2014). pYIN: a fundamental frequency estimator using probabilistic threshold distributions. In *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, pages 659–663.

Mauch, M., Frieler, K., and Dixon, S. (2014). Intonation in unaccompanied singing: Accuracy, drift, and a model of reference pitch memory. *The Journal of the Acoustical Society of America*, 136(1):401–411.

Mendes, A. P., Rothman, H. B., Sapienza, C., and Brown, W. (2003). Effects of Vocal Training on the Acoustic Parameters of the Singing Voice. *Journal of Voice*, 17(4):529–543.

Miller, R. (1996). *On the Art of Singing*. Oxford University Press, USA.

Mithen, S. J. (2007). *The Singing Neanderthal: A Search for the Origins of Art, Religion, and Science*. Harvard University Press, Cambridge, MA. esp. Ch. 16, pp. 246–265.

Morrell, M. J., Harte, C. A., and Reiss, J. D. (2011). Queen Mary's 'Media and Arts Technology Studios' audio system design. In *Audio Engineering Society Convention 130*. Audio Engineering Society.

Mullen, P. (2000). Pitch drift as a result of just intonation. *Journal of the Acoustical Society of America*, 108(5):2618. Abstract presented at the 140th meeting of the Acoustical Society of America.

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

Müllensiefen, D., Gingras, B., and Stewart, L. (2011). Piloting a new measure of musicality: The Goldsmiths' Musical Sophistication Index. Technical report, Goldsmiths, University of London.

Mürbe, D., Pabst, F., Hofmann, G., and Sundberg, J. (2002). Significance of Auditory and Kinesthetic Feedback to Singers' Pitch Control. *Journal of Voice*, 16(1):44–51.

Nichols, B. E. (2016). Task-based variability in children's singing accuracy. *J. Res. Music Educ.*, 64(3):309–321.

Nichols, B. E. and Wang, S. (2016). The effect of repeated attempts and test-retest reliability in children's singing accuracy. *Music. Sci.*, 20(4):551–562.

Pfordresher, P. Q. (2012). Musical training and the role of auditory feedback during performance. *Annals of the New York Academy of Sciences*, 1252(1):171–178.

Pfordresher, P. Q. and Brown, S. (2007). Poor-pitch singing in the absence of 'tone deafness'. *Music Percept.*, 25(2):95–115.

Pfordresher, P. Q., Brown, S., Meier, K. M., Belyk, M., and Liotti, M. (2010). Imprecise singing is widespread. *The Journal of the Acoustical Society of America*, 128(4):2182–2190.

Pfordresher, P. Q. and Mantell, J. T. (2014). Singing with yourself: Evidence for an inverse modeling account of poor-pitch singing. *Cognitive Psychology*, 70:31–57.

Prame, E. (1994). Measurements of the vibrato rate of ten singers. *The Journal of the Acoustical Society of America*, 96(4):1979–1984.

Prout, E. (2011). *Harmony: Its Theory and Practice*. Cambridge University Press.

R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Schoenberg, A. (1978). Theory of harmony. *Roy E. Carter (Berkeley: University of California Press, 1978)*, 25.

Seashore, C. E. (1931). The natural history of the vibrato. *Proceedings of the National Academy of Sciences*, 17(12):623–626.

Seaton, R., Pim, D., and Sharp, D. (2013). Pitch drift in a cappella choral singing. *Proc. Inst. Acoust. Ann. Spring Conf.*, 35(1):358–364.

Stewart, L., von Kriegstein, K., Warren, J. D., and Griffiths, T. D. (2006). Music and the brain: Disorders of musical listening. *Brain*, 129(10):2533–2553.

Sundberg, J. (1977). The acoustics of the singing voice. *Scientific American*, 236(3):82–91.

Sundberg, J. (1987). *The Science of the Singing Voice.* Northern Illinois University Press, DeKalb, IL.

Sundberg, J. (1994). Acoustic and Psychoacoustic Aspects of Vocal Vibrato. Technical Report STL-QPSR 35 (2–3), pages 45–68, Department for Speech, Music and Hearing, KTH.

Sundberg, J., Lã, F. M., and Himonides, E. (2013). Intonation and expressivity: A single case study of classical western singing. *Journal of Voice*, 27(3):391–e1.

Swannell, J. (1992). *The Oxford Modern English Dictionary.* Oxford University Press, USA. p. 560.

Takeuchi, A. H. and Hulse, S. H. (1993). Absolute pitch. *Psychol. Bull.*, 113(2):345.

Terasawa, H. (2004). Pitch Drift in Choral Music. *Music 221A final paper, Center for Computer Research in Music and Acoustics, at Stanford University, CA. Available at https://ccrma. stanford. edu/ hiroko/pitchdrift/paper221A. pdf (Last viewed 5 June 2014).*

Titze, I. R. and Martin, D. W. (1998). Principles of voice production.

Traunmüller, H. and Eriksson, A. (1995). The frequency range of the voice fundamental in the speech of male and female adults. *Consulté le*, 12(02):2013.

Walker, K. M., Bizley, J. K., King, A. J., and Schnupp, J. W. (2011). Cortical encoding of pitch: Recent results and open questions. *Hearing research*, 271(1-2):74–87.

Welch, G. F. (1979). Poor pitch singing: A review of the literature. *Psychology of Music*, 7(1):50–58.

Welch, G. F. (2005). Singing as communication. *Musical Communication*, pages 239–259.

Welch, G. F., Sergeant, D. C., and White, P. J. (1997). Age, sex, and vocal task as factors in singing 'in tune' during the first years of schooling. *Bull. Counc. Res. Music Educ.*, 133:153–160.

Winter, B. (2013). Linear models and linear mixed effects models in r with linguistic applications. *arXiv preprint arXiv:1308.5499.*

Xu, Y. and Sun, X. (2000). How fast can we really change pitch? maximum speed of pitch change revisited. In *INTERSPEECH*, pages 666–669.

Zarate, J. M. and Zatorre, R. J. (2008). Experience-dependent neural substrates involved in vocal pitch regulation during singing. *Neuroimage*, 40(4):1871–1887.