Intonation Trajectories within Tones in Unaccompanied Soprano, Alto, Tenor, Bass Quartet Singing

Jiajie Dai[1, a)] and Simon Dixon[1, b]
Centre for Digital Music, Queen Mary University of London

(Dated: 9 September 2019)

Unlike fixed-pitch instruments, the voice requires careful regulation during each note in order to maintain a steady pitch. Previous studies 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 paper is to study pitch variation within vocal notes and ascertain what factors influence the various parts of a note. We recorded five SATB quartets singing two pieces of music in three different listening conditions, according to whether the singers can hear the other participants or not. After analysing all of the individual notes and extracting pitch over time, we observed the following regularities: 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 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); 5) The listening condition had no significant effect on within-note pitch trajectories.

©*2019 Acoustical Society of America.* [http://dx.doi.org(DOI number)]

[XYZ] Pages: 1–11

PACS numbers: 43.75.Rs, 43.75.Bc, 43.75.Xz, 43.75.St

I. INTRODUCTION

Singing is important because it is the most universal form of music-making (Brown, 1991), and it allows for personal and expressive communication. Unlike external instruments which are mastered by a small minority, almost everyone uses their voice on a daily basis, and can, to some extent, sing. Although singing is common to all human societies and we all have our own idea of what singing actually is (Potter, 2000b, p. 1), many aspects of singing have not been explored in the research literature. For example, the shapes of, and factors that affect, vocal pitch trajectories within notes have yet to be explained. The motivation of this paper is to determine whether pitch trajectories share common shapes, and what factors influence the transient parts of notes.

Intonation, defined as the accuracy of pitch in playing or singing (Swannell, 1992), is regarded as an important aspect of music performance (Sundberg *et al.*, 2013). Such a definition assumes that a reference exists for pitch, and presumably that this reference is fixed at least for each note, enabling accuracy to be assessed, either continuously over the duration of a note, or once for the entirety of a note. For the latter case, previous studies calculated the mean or median of fundamental frequency (f_0) estimates on short audio frames (Howard, 2007; Mauch *et al.*, 2014). For analysis of intonation within a note, the frame-level estimates describe the pitch trajectory as a time series.

The complexity of the vocal apparatus makes it difficult to sing accurately. Voice production requires the coordination of the lungs, vocal folds, larynx, pharynx and mouth (Sundberg, 1977). To produce a tone at a given pitch also requires muscle memory and tonal memory (Alldahl, 2008). Most people, who do not have perfect pitch (the ability to recognise the pitch of a note or produce any given note), rely on a recent reference for intonation (Takeuchi and Hulse, 1993). Therefore, the instrumental accompaniment or reference pitch is crucial for the tuning.

Previous studies have explored vocal pitch trajectories for singing voice synthesis, especially for performance modelling (Umbert *et al.*, 2015), and modelled the observed pitch in an imitation task, given a time-varying stimulus pitch (Dai and Dixon, 2016). This paper presents an exploratory study to find which factors have an effect on the pitch trajectory of vocal notes. There are many factors influencing overall intonation accuracy, such as score information (e.g. target pitch, duration, intervals between the target and simultaneous or recent

[a)] j.dai@qmul.ac.uk
[b)] s.e.dixon@qmul.ac.uk

pitches), individual differences (e.g. sex, training background), and the accompaniment.

We created a public data set which involves 20 participants (five groups of four) singing two pieces of music in three different listening conditions: solo, with one vocal part missing and with all vocal parts. Each participant sings their usual vocal part: soprano, alto, tenor or bass. SATB singing was chosen for this intonation study as it is a common configuration for singing ensembles in Western music.

The remainder of the paper is structured as follows. Section II discusses existing work related to singing intonation and interaction. Section III contains our research questions, experimental design and methodology. In Section IV, we describe our data analysis, including annotation and calculation of intonation metrics. Section V presents our results, which are then discussed in Section VI. Our conclusions are found in Section VII, followed by details of where the annotated data and software can be freely obtained in Section VIII.

II. PREVIOUS WORK

Intonation accuracy is one of the features used to evaluate a singer's performance. For calculating intonation accuracy from an audio recording, pitch and fundamental frequency (f_o) are generally treated as exchangeable (see Section IV A). In the 1930s, Seashore measured fundamental frequency in recordings of renowned singers and revealed considerable departures from equally tempered tuning (Seashore, 1914; Sundberg *et al.*, 2013). Since that time, many studies on singing and intonation focus on accuracy, especially measuring the pitch error, which is the difference between the observed pitch and a predetermined target pitch. Some studies investigate the pitch drift of singing ensembles (e.g. Devaney and Ellis, 2008; Howard, 2003; Kalin, 2005; Terasawa, 2004) or solo singers (Mauch *et al.*, 2014). Other studies investigate factors that influence the pitch error (e.g. Pfordresher *et al.*, 2010; Welch *et al.*, 1997). In a previous study (Dai and Dixon, 2017), we observed that pitch error and melodic interval error increase when singers can hear each other, and in particular, that singing without the bass part had less mean absolute pitch error than singing with all vocal parts. In addition, we found that pitch variation within notes was lower when participants sang solo than with their partners.

Besides pitch error, other studies have investigated interval error, the extent to which pitch differences between subsequent (melodic interval error) or simultaneous (harmonic interval error) tones deviate from their target values. Some melodic intervals were reported as being harder to sing than other intervals, such as tritones (Dai *et al.*, 2015) and perfect fifths (Vurma and Ross, 2006). There is a phenomenon called *compression*, whereby sung melodic intervals tend to be smaller than the target intervals (Pfordresher and Brown, 2007). Harmonic intervals constitute another important factor which influences intonation. Hagerman and Sundberg (1980) studied the harmonic intervals sung by two barbershop quartets, and found that intervals did not reflect just or Pythagorean tuning as expected, although the singing was precise (low standard deviations). They suggested that deviations from pure intervals (i.e. where the frequencies of notes are related by ratios of small whole numbers (Lindley, 2001)) could be due to aperiodicity in the voice, which broadens the spectral peaks and renders beats inaudible. They also observed a general stretching of intervals in performance, which they describe as sounding "more active and expressive than flat intervals". Nordmark and Ternström (1996) investigated the preferred tuning of major third intervals, finding that participants tuned intervals closer to equal temperament than pure intonation. Howard (2007) observed the use of non-equal-tempered tuning in unaccompanied singing, although his data did not fully confirm his predictions based on the use of pure intervals. In both cases singers produced intervals between the pure and equal-tempered versions of the intervals, while Devaney *et al.* (2012) observed that some intervals were close to either just or Pythagorean tunings, but most were within a standard deviation of equal-temperament.

For an individual singer, singing is a complicated task involving both perception and production. The voice organ can be viewed as an instrument consisting of a power supply (the lungs), an oscillator (the vocal folds) and a resonator (the larynx, pharynx, and mouth) (Sundberg, 1977). 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). External influences such as reference pitches provided by accompaniment also affect pitch accuracy.

Interaction is an important factor for ensemble singing, which is a cooperative activity involving communication within the ensemble and with the audience (Potter, 2000a, p. 158). Few people can produce a correct pitch directly without the use of an external reference pitch (Takeuchi and Hulse, 1993), such as that provided by instrumental accompaniment. Although accompaniment has been shown to enhance the individual learning of a piece (Brandler and Peynircioglu, 2015), it can also reduce pitch accuracy during singing, even when the accompaniment is in unison with the singer (Dai and Dixon, 2016, 2017; Pfordresher and Brown, 2007). Most singers adjust their intonation using auditory feedback to reach the intended note (Zarate and Zatorre, 2008), and accompaniment might distract singers from hearing their own feedback.

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. Alldahl, 2008; Bohrer, 2002). Dai and Dixon (2017) observed that pitch error and melodic interval error increase when the participants can hear other singers, but harmonic interval error is reduced when all singers

hear each other. In unaccompanied multi-part singing, Howard (2007) demonstrated how singing pure intervals can cause drift, and he found that singers do in fact tend to non-equal-tempered tuning and drift in pitch with modulation. With different musical material, Devaney *et al.* (2012) observed that only some intervals were significantly different from equal temperament; in their study, the singers did not exhibit a large amount of drift.

Individual differences such as age and sex also influence pitch accuracy (Welch *et al.*, 1997). Likewise, musical training and experience have some influence; Mauch *et al.* (2014) found that self-rated singing ability and choir experience, but not general musical background, correlated significantly with intonation accuracy. Singers who exhibit much greater than average pitch errors are classified as *poor singers*, a phenomenon that has been the focus of several studies (Berkowska and Dalla Bella, 2009; Dalla Bella *et al.*, 2007; Pfordresher and Brown, 2007; Pfordresher *et al.*, 2010).

Observation of the pitch trajectory within individual notes reveals transient parts at the beginning and end of each note. At the beginning of a tone, a pitch glide is often observed as the singer adjusts the vocal cords 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; Seashore, 1931; Sundberg, 1995). 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 cord state) if there is no immediately following note (Xu and Sun, 2000). Pitch variation within a note has been modelled for vocal synthesis, as well as note level features (onset and offset), intra- and internote features (changes within and between notes), and the relationship to timbre variations (Umbert *et al.*, 2015).

Vibrato is used to add expression to vocal and instrumental music. In singing, it can occur spontaneously through variations in the larynx. Professional (particularly opera) singers tend to produce vibrato: a periodic modulation of $\mathrm{f_o}$, which is not normally used in speech (Sundberg, 1987). The frequency of the vibrato is usually in the range 5–8 Hz according to the vibrato type (Fischer, 1993). Although all human voices can produce vibrato, it has been shown that with training, singers are able to elicit control over both vibrato rate and depth (Dromey *et al.*, 2003; King and Horii, 1993).

Several software systems for pitch analysis have been developed which support scientific measurement, such as Praat (Boersma, 2002), Sound Visualiser (Cannam *et al.*, 2006) and Tony (Mauch *et al.*, 2015). de Cheveigné and Kawahara (2002) introduced a pitch extraction method, YIN, which has been applied extensively. This algorithm improves upon the autocorrelation method by means of a difference function plus several modifications that improve system performance. PYIN (Mauch and Dixon, 2014) is a probabilistic extension of YIN which enhances robustness against errors and is employed in the Tony software used in this paper.

III. METHODOLOGY

In this section, we describe our exploratory research questions, the experimental design, musical material, participants and experimental procedure. Links to the data and score information can be found in Section VIII.

A. Listening condition

For our experiment, three *listening conditions* were defined based on what the singer can hear as they sing. In the *closed condition*, the singer can only hear their own voice and metronome, thus they are effectively singing solo. In the *partial condition*, the singer can hear some, but not all of the other vocal parts. This is achieved by physically 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 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 1 visualises the listening and test conditions.

B. Research questions

This study of interactive intonation in unaccompanied SATB singing is driven by a number of research questions. Firstly, we wish to know whether there are patterns or regularities in the pitch trajectories of individual notes. We expect 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 questions is which factors influence the tendencies of the transient part. The note trajectories might show significant differences due to context, such as when singing after a higher pitch or a lower pitch. We also wish to determine whether pitch trajectories differ by vocal part or sex. We previously observed significant differences between vocal

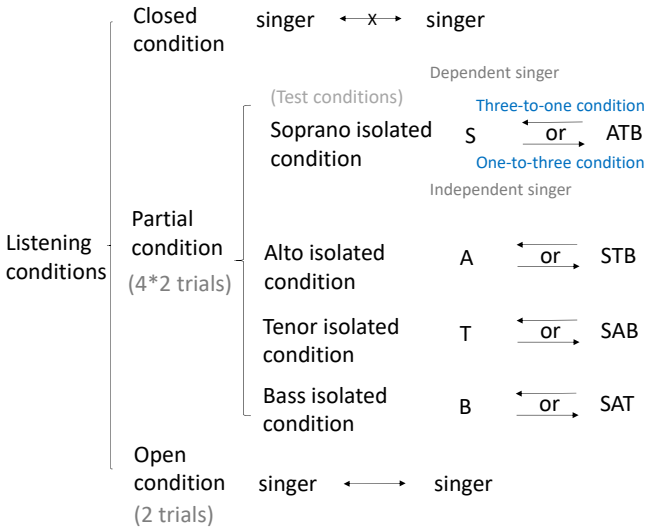

FIG. 1. Listening and test conditions. The arrows indicate the direction of acoustic feedback.

parts in terms of pitch error (Dai, 2019; Dai and Dixon, 2017, 2019b). Finally, we would like to see 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?

C. 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 paper.) The age range was from 20 to 55 years old (mean: 28.0, median: 26.5, std.dev.: 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 (Müllensiefen *et al.*, 2014) to test the effect of training. The participants had an average of 3.3 years of music lessons and 5.8 years of singing experience.

D. 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 (see Section IV B): the first piece has 34 unique harmonic intervals between parts and the second piece has 30 harmonic intervals. To control the duration of the experiment, we shortened the original score by deleting the repeats. We also reduced the tempo 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 VIII.

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).

E. Procedure

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 self-identified voice type, 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) × 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. We use the performance of the isolated singer in the one-to-three condition 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.

IV. DATA ANALYSIS

This section describes the annotation procedure and the measurement of pitch error and harmonic interval error. The experimental data comprises 5 (groups) × 4 (singers) × 2 (pieces) × 10 (trials) = 400 audio files, each containing 65 to 116 notes. Any missing notes were excluded from the analysis. The software *Tony* (Mauch *et al.*, 2015) was chosen as the annotation tool. *Tony* performs pitch detection using the pYIN algorithm, which outperforms the YIN algorithm (Mauch and Dixon, 2014), and then automatically segments pitch trajectories into note objects, and provides a convenient interface for manual checking and correction of the resulting annotations. The automatic segmentation, based on note energy and pitch changes, provided the note onset and offset times for our data, and rarely needed any correction.

For each audio file, we exported two `.csv` files, one containing the note-level information (for calculating pitch and interval errors) and the other containing the pitch trajectories. It took about 67 hours to manually check and correct the 400 files, resulting in 37246 annotated pitch values, which were stored with metadata on the singer, experimental condition and score. The information in our database includes: group number, singer number, vocal part, 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, normalised note trajectories, real-time note trajectories, age, sex and questionnaire scores. MATLAB 2015a was used for statistics and modelling.

A. Conversion of f_o

The *Tony* software segments the recording into notes and silences, and outputs the median fundamental frequency f_o for each note, as well as the f_o value for each 5.8 ms frame. The conversion of fundamental frequency to musical pitch p is calculated as follows:

$$p = 69 + 12\log_2\frac{f_o}{440}. \tag{1}$$

This scale is chosen such that its units are semitones (one semitone is equal to 100 cents), 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.

B. Intonation Metrics

Intonation accuracy is quantified in terms of pitch error and harmonic interval error, as defined below. 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). This is usually defined on the level of notes, but can also be measured for each sampling point of the pitch trajectory:

$$e_i^p = p_i^o - p_i^s \tag{2}$$

where p_i^o is the observed pitch in a single frame of note i (or the median $\bar{p}_i$ over the duration of the note), and p_i^s is the score pitch of note i.

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

$$\text{MAPE} = \frac{1}{M}\sum_{i=1}^{M}|e_i^p| \tag{3}$$

A musical *interval* is the difference between two pitches (Prout, 2011), which is proportional to the logarithm of the ratio of the fundamental frequencies of the two pitches. We distinguish two types of interval: *melodic intervals*, where the two notes are sounded in succession; and *harmonic intervals*, where both notes sound simultaneously (although they might not start simultaneously). In this paper we consider only the harmonic interval error, defined as the difference between the observed and score intervals:

$$e^h_{i,A,j,B} = (\bar{p}_{i,A} - \bar{p}_{j,B}) - (p^s_{i,A} - p^s_{j,B}) \tag{4}$$

where $p^s_{i,A}$ and $p^s_{j,B}$ are the score pitches of two overlapping 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 are not assumed to be equal.

V. 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). We excluded from the results any notes which had a duration shorter than 0.15 seconds (4.1%) or MAPE larger than one semitone (12.0%).

A. The shape of note trajectories

To observe regularities in note trajectories across differing note durations, we compared two methods of equalising the time-scale of trajectories: normalisation and truncation. Normalised pitch trajectories are expressed as a function of the fraction of the note that

has elapsed (from 0 to 1), while for the truncated trajectories, the beginning and end of the note are modelled separately, using respectively the first and last 0.4 seconds of the note (77% of notes are over 0.4 seconds, and 55% over 0.55 seconds in duration). For comparing trajectories of different score pitches, we use the pitch error, that is, the deviation from the target (score) pitch.

For the normalisation method, the note trajectories were re-sampled to 100 sampling points with the MATLAB resample function. Then any common shape of vocal notes can be obtained by averaging across notes. Figure 2 plots the resulting note trajectory generated by calculating the mean of all the sampling points. For comparison, we also show the absolute pitch error, which is much larger in magnitude.

In Figure 2 we observe transient parts at the beginning and end of the note. Based on the slope of the MPE curve, the initial and final transients each comprise about 15-20% of the note's duration. In the following, we take the first 15% and the final 15% of each note as the transient parts. The length of the two transient parts is approximately the same, and the shape is almost symmetrical, consisting of peaks at both ends of the note, with a relatively stable middle portion. The mean pitch error is negative, reflecting a tendency to sing flat relative to the score pitch.

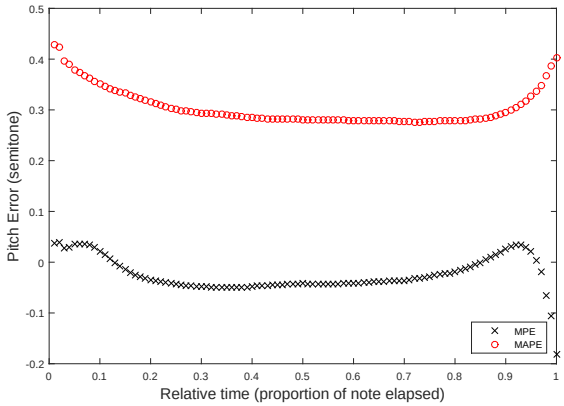

FIG. 2. Average pitch trajectory within a tone expressed as mean pitch error (MPE) across time-normalised notes, with mean absolute pitch error (MAPE) shown for comparison.

An alternative way to combine note trajectories of varying length is to 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 3 and 4. From these figures, we observe that the first 0.12 and last 0.12 seconds of each note have the most pitch variance. This corresponds to about 15-20% of the mean note duration (0.69 seconds). This result is similar to that for normalised trajectories (Figure 2), where the initial sharp fall and final rise in pitch are not as sharp due to the normalisation of different length notes. The average results hide differences in the proportion and direction of transients which arise due to individual differences, score pitch and vocal part, which will be investigated in the following sections.

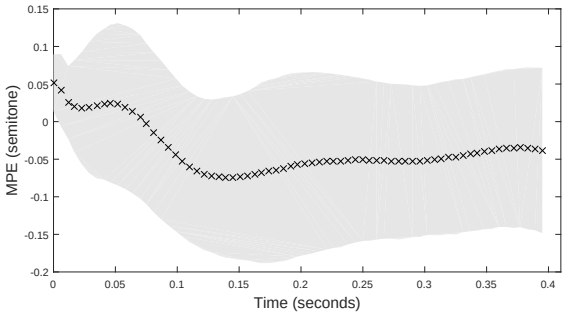

FIG. 3. Mean pitch error (crosses) and range of one standard deviation from the mean (shaded) for the initial 0.4 seconds of each note

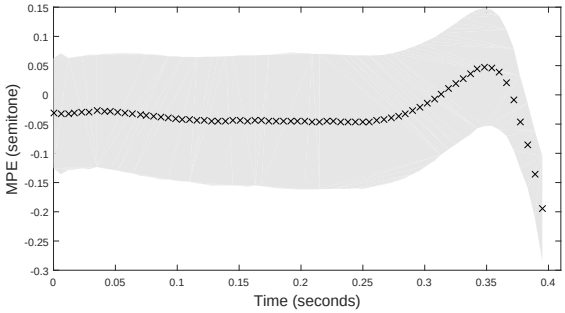

FIG. 4. Mean pitch error (crosses) and range of one standard deviation from the mean (shaded) for the final 0.4 seconds of each note

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

In Figure 3, 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 we now examine.

B. Adjacent pitch

In the previous section, we observed large pitch fluctuations at each end of the note. To test whether these

fluctuations are influenced by adjacent pitches in the score, we separate the data for each end of a note into two situations, based on whether the previous (respectively next) pitch is lower or higher than the current pitch. Repeated pitches are ignored. 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. In Figure 5 we observe that singers tend to overshoot the target pitch and then adjust downward after singing a lower pitch, while after a high pitch they reach the target almost immediately. Jers and Ternström (2005, Fig. 3-4) observed that singers also overshoot the interval (undershoot the pitch) before correcting when they transition from a higher pitch to a lower pitch. The steady state pitch is 1 cent sharper when coming from a lower pitch than when the previous pitch is higher ($F(1, 38) = 77.97, p < 0.001$). 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). 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.

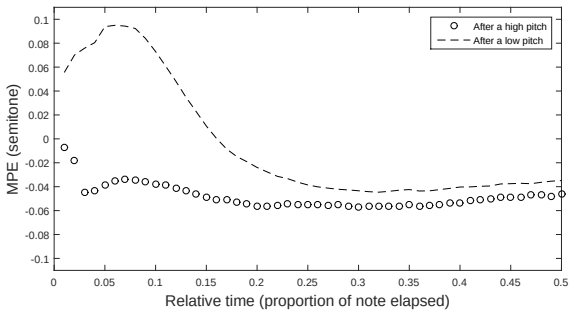

FIG. 5. The effect of singing after a lower or higher pitch: mean pitch error in relative time

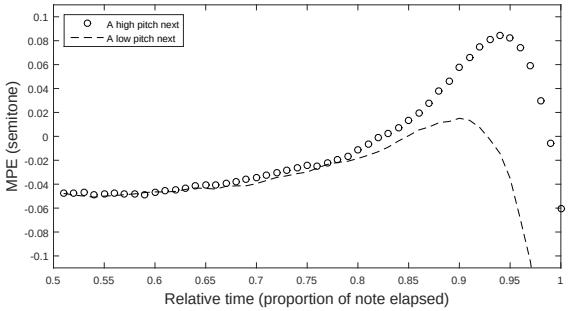

FIG. 6. The effect of singing before a lower or higher pitch: mean pitch error in relative time

This phenomenon is also observed between sexes. An ANOVA shows a significant difference between the note beginnings of male singers and female singers ($F(3, 76) = 59.37, p < .001$). In general, male participants sing 11 cents flatter while females sing 4 cents sharper than the score pitch. 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.

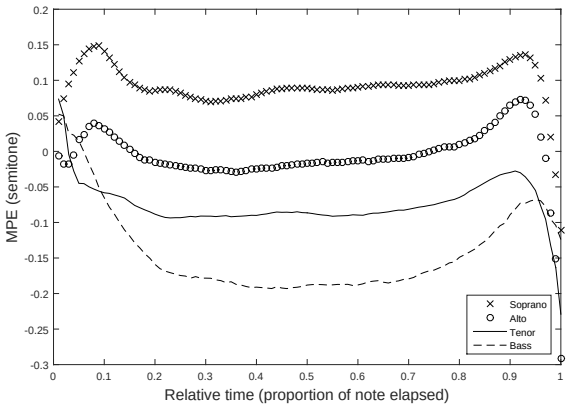

FIG. 7. Mean pitch error over note duration for each vocal part

C. Vocal parts and sex

To explore the factor of vocal part, the normalised note trajectories were plotted for each of the four vocal parts (Figure 7). Firstly, we observe about an 8-cent pitch difference between each pair of adjacent parts in our 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).

D. Modelling the note trajectories

For a better understanding of the tendencies of pitch trajectories, we modelled them as three separate components: initial transient, note middle and final transient. As discussed previously, the transient parts were defined by the first 15% and last 15% of the duration. The tendency of each component was approximated by linear regression. Figure 8 shows an example of a single pitch

trajectory and the linear fits for each of the three components.

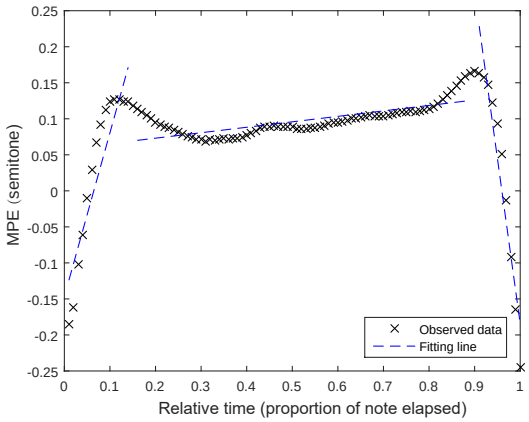

FIG. 8. Example of the pitch trajectory of a single note and the fitting lines for the initial, middle and final components of the note

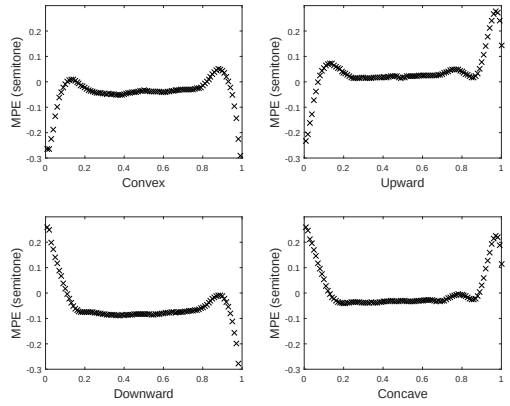

FIG. 9. Mean pitch trajectories of the four trajectory classes in relative time (proportion of note elapsed)

To describe the different types of trajectories, we classify them into four categories (Concave, Convex, Upward, Downward) according to the slopes of their initial and final transients, which are either positive or negative. Table I shows that the most popular shapes are Convex and Downward, both of which have a negative note release.

Table II 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. Although the initial and final slopes have high variance, Figures 3 and 4 show that their starting and ending points are less spread (smaller standard deviation) than the rest of the trajectory.

E. Listening condition

Finally, the influence of listening condition on note trajectory classification was considered. The influence of listening condition on mean pitch is discussed in previous work (Dai and Dixon, 2017, 2019a). In this paper, the ANOVA test on the note trajectories did not show any significant difference between listening conditions (solo, partial independent, partial dependent and open) for initial transient ($F(3, 49194) = 0.07, p = 0.79$), middle section ($F(3, 49194) = 1.62, p = 0.20$), and final transient ($F(3, 49194) = 0.05, p = 0.83$).

VI. DISCUSSION

In this study, we observed 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. Unlike the piano, where stretching of intervals is related to the inharmonicity of the partials, sung tones are not inharmonic, so we are unable to explain this observation. Hagerman and Sundberg (1980) also observed stretched harmonic intervals in an experiment with barbershop singers, giving the explanation that they sound "more active and expressive".

If we compare 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 (Devaney and Ellis, 2008; Howard, 2003; Terasawa, 2004), and is typically downward in direction, although upward drift has also been observed.

While the averaged note trajectories, particularly when sorted into categories (Figure 9), show quite smooth curves, the individual pitch trajectories exhibit much greater degrees of variation (e.g. Figure 8, 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, in Figures 2 and 3 we observe a concave shape (a small local minimum) in the first 5% (respectively 0.04 seconds) of the note trajectory. If we compare with Figure 7, 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 I).

The differences observed in the averaged curves are small in magnitude, of the same order as the just noticeable difference in pitch (about 5 cents, Loeffler (2006)) .

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 I. 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

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

Many of the sung examples have larger differences, which are reduced by the averaging process, but are likely to be perceptible in the original examples. A listening test using synthetic stimuli would be required to identify the perceptual relevance of the features of pitch trajectories identified in this paper.

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 9).

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 our results significantly if they were to be segmented differently.

Although vibrato is a feature of many singing pitch trajectories, we did not explicitly model it in this work (cf. Dai and Dixon, 2016; Mehrabi *et al.*, 2017). 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 we did not observe strong vibrato in our data, and in the cases where vibrato was present, it tended to be uneven, which would make it difficult to model.

VII. CONCLUSIONS

In this paper, we present a study of pitch trajectories of single notes in multi-part singing. According to our analysis of over 35000 individual notes, we find a general shape of vocal notes 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, or 15-20% of note duration. 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. We also observe 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.

In conclusion, the main contribution of this paper is the observation, measurement and analysis of the note transient parts by characterising their shapes and influencing factors. Although many further issues remain to be investigated, we hope that the current observations provide a better understanding of the singing voice.

VIII. DATA AVAILABILITY

The code and the data needed to reproduce our 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

ACKNOWLEDGMENTS

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

Many thanks to all of the participants who contributed to this project. We also thank Marcus Pearce, Daniel Stowell and Christophe Rhodes for their advice on experimental design. Jiajie Dai is supported by a China Scholarship Council and Queen Mary Joint PhD Scholarship.

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

Berkowska, M., and Dalla Bella, S. (**2009**). "Acquired and congenital disorders of sung performance: A review," Adv. Cogn. Psychol. **5**(-1), 69–83, doi: 10.2478/v10053-008-0068-2.

Boersma, P. (**2002**). "Praat, a system for doing phonetics by computer," Glot Int. **5**(9/10), 341–345.

Bohrer, J. C. S. (**2002**). "Intonational strategies in ensemble singing," Ph.D. 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," J. Res. Music Educ. **63**(3), 281–297.

Brown, D. (**1991**). *Human Universals* (Temple University Press, Philadelphia), pp. 1–160.

Cannam, C., Landone, C., Sandler, M. B., and Bello, J. P. (**2006**). "The Sonic Visualiser: A visualisation platform for semantic descriptors from musical signals," in *7th Int. Conf. Music Inf. Retr.*, pp. 324–327.

Dai, J. (**2019**). "Modelling intonation and interaction in vocal ensembles," Ph.D. thesis, Queen Mary University of London, pp. 104–115.

Dai, J., and Dixon, S. (**2016**). "Analysis of vocal imitations of pitch trajectories," in *17th Int. Soc. Music Inf. Retr. Conf.*, pp. 87–93.

Dai, J., and Dixon, S. (**2017**). "Analysis of interactive intonation in unaccompanied SATB ensembles," in *18th Int. Soc. Music Inf. Retr. Conf.*, pp. 599–605.

Dai, J., and Dixon, S. (**2019**a). "Singing together: Pitch accuracy and interaction in unaccompanied unison and duet singing," J. Acoust. Soc. Am. **145**(2), 663–675.

Dai, J., and Dixon, S. (**2019**b). "Understanding intonation trajectories and patterns of vocal notes," in *Int. Conf. Multimedia Modell.*, Springer, pp. 243–253.

Dai, J., Mauch, M., and Dixon, S. (**2015**). "Analysis of intonation trajectories in solo singing," in *16th Int. Soc. Music Inf. Retr. Conf.*, pp. 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, doi: 10.1121/1.2427111.

de Cheveigné, A., and Kawahara, H. (**2002**). "YIN, a fundamental frequency estimator for speech and music," J. Acoust. Soc. Am. **111**(4), 1917–1930, doi: 10.1121/1.1458024.

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

Devaney, J., Mandel, M. I., and Fujinaga, I. (**2012**). "A study of intonation in three-part singing using the automatic music performance analysis and comparison toolkit (AMPACT)," in *13th Int. Soc. Music Inf. Retr. Conf.*, pp. 511–516.

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

Fischer, P.-M. (**1993**). *Die Stimme des Sängers: Analyse ihrer Funktion und Leistung-Geschichte und Methodik der Stimmbildung (The Voice of the Singer: Analysis of Function and Performance - History and Methodology of Voice Training)* (Metzler, Stuttgart, Germany).

Gerhard, D. (**2005**). "Pitch track target deviation in natural singing," in *6th Int. Conf. Music Inf. Retr.*, pp. 514–519.

Hagerman, B., and Sundberg, J. (**1980**). "Fundamental frequency adjustment in barbershop singing," J. Res. Singing **4**, 1–17.

Howard, D. M. (**2003**). "A capella SATB quartet in-tune singing: Evidence of intonation shift," in *Stockholm Music Acoust. Conf.*, Vol. 2, pp. 462–466.

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

Jers, H., and Ternström, S. (**2005**). "Intonation analysis of a multichannel choir recording," Speech Music Hearing Q. Prog. Status Rep. **47**(1), 1–6.

Kalin, G. (**2005**). "Formant frequency adjustment in barbershop quartet singing," Master's thesis, Royal Institute of Technology, Dept of Speech Music and Hearing, Stockholm.

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

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," Ph.D. thesis, Georgia Institute of Technology.

Mauch, M., Cannam, C., Bittner, R., Fazekas, G., Salamon, J., Dai, J., Bello, J., and Dixon, S. (**2015**). "Computer-aided melody note transcription using the Tony software: Accuracy and efficiency," in *Proc. 1st Int. Conf. Technol. Music Not. Represent.*, pp. 23–30.

Mauch, M., and Dixon, S. (**2014**). "PYIN: A fundamental frequency estimator using probabilistic threshold distributions," in *IEEE Int. Conf. Acoust., Speech, Signal Process.*, pp. 659–663.

Mauch, M., Frieler, K., and Dixon, S. (**2014**). "Intonation in unaccompanied singing: Accuracy, drift, and a model of reference pitch memory," J. Acoust. Soc. Am. **136**(1), 401–411.

Mehrabi, A., Dixon, S., and Sandler, M. (**2017**). "Vocal imitation of synthesised sounds varying in pitch, loudness and spectral centroid," J. Acoust. Soc. Am. **143**(2), 783–796.

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

Nordmark, J., and Ternström, S. (**1996**). "Intonation preferences for major thirds with non-beating ensemble sounds," Speech Music Hearing Q. Prog. Status Rep. **37**(1), 57–62.

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," J. Acoust. Soc. Am. **128**(4), 2182–2190.

Potter, J. (**2000**a). "Ensemble singing," in *The Cambridge Companion to Singing*, edited by J. Potter (Cambridge Univer-

sity Press, Cambridge, UK), pp. 158–164, doi: 10.1017/CCOL9780521622257.014.
Potter, J. (**2000**b). "Introduction: singing at the turn of the century," in *The Cambridge Companion to Singing*, edited by J. Potter (Cambridge University Press, Cambridge, UK), pp. 1–5, doi: 10.1017/CCOL9780521622257.001.
Prout, E. (**2011**). *Harmony: Its Theory and Practice* (Cambridge University Press).
Seashore, C. E. (**1914**). "The tonoscope," Psychol. Monogr. **16**(3), 1–12.
Seashore, C. E. (**1931**). "The natural history of the vibrato," Proc. Natl Acad. Sci. **17**(12), 623–626.
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," Sci. Am. **236**(3), 82–91.
Sundberg, J. (**1987**). *The Science of the Singing Voice* (Northern Illinois University Press, DeKalb, IL).
Sundberg, J. (**1995**). "Acoustic and psychoacoustic aspects of vocal vibrato," in *Vibrato*, edited by P. Dejonckere, M. Hirano, and J. Sundberg (Singular Publishing Group, San Diego, CA), pp. 35–62.
Sundberg, J., Lã, F. M., and Himonides, E. (**2013**). "Intonation and expressivity: A single case study of classical Western singing," J. 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, Stanford University, URL https://ccrma.stanford.edu/~hiroko/pitchdrift/paper221A.pdf.
Umbert, M., Bonada, J., Goto, M., Nakano, T., and Sundberg, J. (**2015**). "Expression control in singing voice synthesis: Features, approaches, evaluation, and challenges," IEEE Signal Process. Mag. **32**(6), 55–73.
Vurma, A., and Ross, J. (**2006**). "Production and perception of musical intervals," Music Percept. **23**(4), 331–344.
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.
Xu, Y., and Sun, X. (**2000**). "How Fast Can We Really Change Pitch? Maximum Speed of Pitch Change Revisited.," in *6th Int. Conf. Spoken Lang. Process.*, pp. 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.