Modelling Intonation Trajectories and Understanding the Pattern of Singing Notes *

First Author[1[0000−1111−2222−3333]], Second Author[2,3[1111−2222−3333−4444]], and Third Author[3[2222−−3333−4444−5555]]

No Institute Given

Abstract. Different from instruments hold the same pitch over time, singing an accurate pitch is a problem that all singers have to face. Previous studies have investigated singing performance which takes single note as an element, such as intonation accuracy, pitch drift while the note trajectory inside the notes hardly been investigated. The aim of this project is to find the pattern internal the vocal notes and figure out what factors may influence the pitch various. In particular, we recorded data which including five SATB groups (four participants as a group) sang two pieces of music in ten repetitions. After extracted fundamental frequency and annotation, we obtained all the notes by unit duration and real-time duration, then observed a regular pattern among all the notes. To be specific: there is a transient part in both the note starting and ending position which the beginning transient is about 15-20%, the ending transient is about 10-20% at the end. The shape of transient parts is significantly different while the amplitude of the transient is according to adjacent pitch. The results may benefit vocal synthesis and guide musical education.

Keywords: Intonation accuracy · Note pattern · Pitch trajectories.

1 Introduction

Singing moves and excites people, which directly express our personalities and emotions. Different from instruments which have hundreds of years of technical development behind it, almost everyone can speak, everyone can sing, so we all have our own idea of what singing actually is [15, p. 1]. Although singing is the most common instrument to all human societies [3], the patterns and factors that affect the vocal trajectories are hardly been explained. The motivation of this paper is trying to explain whether a general pattern of vocal note exists, and how factors of influence effect on the pitch various inside the notes.

Intonation is commonly regarded as an important aspect of music performance [17], which describe how a pitch is played or sang in tune [8], and also describe as the accuracy of pitch in playing or singing [18]. This accuracy could be inside or an entirety of a note. For the entirety intonation accuracy, some

* Supported by organization x.

of the previous studies calculated the mean or median of the fundamental frequency (F0) inside the note (such as [7] and [12]). For the intonation inside a note, some software such as *Tony*[10] could export the pitch information in time series. We measure the intonation as the signed pitch difference compares with the score pitch, labelled in semitones on an equal-tempered scale. The beginning and ending parts of the note trajectories which have more pitch various are call transient parts.

People hardly can produce the correct pitch directly without the use of an external reference pitch [19]. Most singers adjust their intonation using auditory feedback to reach the intended note [21]. Howard tried to prove that just intonation (the frequencies of notes are related by ratios of small whole numbers [9]) causes drift and found out that singers tend to non-equal-tempered tuning and do shift their pitch with modulation [7]. Devaney *et al.* reported that singers tended toward equal temperament and did not exhibit a large amount of drift. These evidences show the insatiability of the human tuning.

Our voice organs which produce speech and singing are extremely complex, which makes human are hard to tune. The voice production requires the cooperation of the lungs, vocal folds, larynx, pharynx and mouth [16]. To produce a voice in a particular pitch also require muscle memory and tonal memory [1]. For most people without the perfect pitch (the ability to recognise the pitch of a note or produce any given note), they tune their intonation rely on a recent reference [19]. Therefore, the instrumental accompaniment or reference pitch is very crucial for the tuning. For the circumstance that singers have to produce the performance without instrumental accompaniment, to understand the mechanism and pattern of the note may benefit the society.

This paper is a exploratory research to To find which factor has a effect on note trajectory. There are many influencing factors which effect on overall intonation accuracy of a note. Such as score information, the individual difference (gender, training background [4]), with or without a accompaniment (instrument or singing ensembles). Some of the factors may have a variable effect, for example, instrumental accompaniment has been shown to enhance the individual learning of a piece [2], it can also reduce pitch accuracy during singing, even when the accompaniment is another singer who sings the exactly the same piece with you [14, 6].

Besides the overall intonation accuracy of an note, the intonation trajectory inside a note is an area need further investment. Different from a fixed pitch instrument, both of the singing and reference pitch can change over time. The previous study has explored the note trajectory for singing voice synthesis, especially for the performance modelling [20]. Or modelling the observed pitch with the stimulus pitch which given to the participants to imitate [5]. The paper focuses on more factors which have an effect on note trajectory and pattern, according to the real-time performance.

2 Methodology

2.1 Research questions

The aim of this paper is to investigate the existence of a pattern for all the note trajectories. If it exist, is there any regulation inside the note. Based on previous research and musical experience, the duration and proportion of the transient parts is important to this project. Whether sing after a high pitch or a low pitch influence the note trajectories? How does this over tune phenomenon happen?

2.2 Participants

20 adult amateur singers (10 male and 10 female) with choir experience volunteered to take part in the study. The age range was from 20 to 55 years old (mean: 27.95, median: 26.50, std.dev.: 7.84). Participants were compensated £10 for their participation. The participants were capable to sing their parts comfortably and they were given the score and sample audio files at least 2 weeks before the experiment. They came from the music society and *a capella* society of the university and a local choir. There was one group of the pilot experiment which participants came from our research group while data did not use for the analysis.

Since training is a crucial factor for intonation accuracy, all the participants were given a questionnaire based on the Goldsmiths Musical Sophistication Index [13] to testing the effect of training. The participants had an average of 3.3 years of music lessons and 5.8 years of singing experience.

2.3 Materials

Two contrasting musical pieces were selected for this study: a Bach chorale, "Oh Thou, of God the Father" and Leo Mathisen's jazz song "To be or not to be".

To control the duration of the experiment, we shortened the original score by deleting the repeat melody. 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.

The equipment included an SSL MADI-AX converter, five cardioid microphones and four loudspeakers. 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).

2.4 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 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 repetitions by each group. 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.

3 Data Analysis

This section describes the annotation procedure and the measurement of pitch error, melodic interval error, harmonic interval error. These metrics of accuracy are applied to the Section 4 for generate and analyse the pattern of note trajectories.

3.1 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* [10] was chosen as the annotation tool. *Tony* performs pitch detection using the PYIN algorithm, which outperforms the YIN algorithm [11], and then automatically segments pitch trajectories into note objects, and provides a convenient interface for manual checking and correction of the resulting annotations. For the missing notes and unrecognised note, we leave the note information blank and excluded them in the analysis.

3.2 Conversion of F0

The *Tony* software segments the recording into notes and silences, and outputs the median fundamental frequency f_0 for each note. The conversion of fundamental frequency to musical pitch p is calculated as follows:

$$p = 69 + 12 \log_2 \frac{f_0}{440}. \qquad (1)$$

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.

3.3 Intonation Metrics

To quantify the effects of interaction on intonation, we measure pitch accuracy in terms of pitch error, melodic interval error, harmonic interval error and note stability, defined below.

Pitch Error Assuming that a reference pitch has been given, *pitch error* can be defined as the difference between observed pitch and score pitch [12]:

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

where $\bar{p}_i$ 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, 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:

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

4 Results

After annotation and half-manually check the segmentation, there are 12300 single notes for each vocal part. Every single note has it own pitch trajectories which is 174 sampling points per seconds (default value from the software *Tony*). The score duration is 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), 89.5% of the score duration in the data is less than one seconds. We excluded the notes which has note duration shorter than 0.15 seconds (4.1%) and MAPE bigger than one semitone (12.0%) for all the results.

To observe the pattern of common note trajectories of all the observed data, two measurements were used according to the note duration: 1) unit note 2) real-time note duration. For the unit note duration, the duration was re-scaled to unit one. For the real-time duration, only the first 0.4 and last 0.4 seconds real time duration were extracted (77% notes have duration longer than 0.4 seconds). All the data is available in .csv file, MATLAB 2015a was used for statistics and modelling.

4.1 The pattern of note trajectories

Because all the collected single notes have different duration and pitch trajectories, it is hard to unify them, the note trajectories were re-sampled to 100 sampling points for an equal size by MATLAB resample function. Then the common shape of the vocal note can be generated as unit one. The unit of note duration (x-axis) was replaced by a number between 0 and 1 note, while the pitch error was used to describe the trajectories (y-axis). After accumulating all the available data, a plot of the trajectory of unit note (Figure 1) was generated by calculate the mean of all the sampling points.

According to the Figure 1 there are two transient parts in the beginning and at the end of the notes which are transient parts. By calculating the differential coefficient, the beginning transient is about 15-20% in the beginning while the ending transient is about 10-20% at the end. The length of two transient parts is approximately the same. We take the 15% in the beginning and the end as the transient parts in the following computing. From the Figure 1 we observe peaks at both the beginning and ending transient parts. As for the middle note, the mean pitch error is smaller than zero means more notes tend to sing flat than the score pitch.

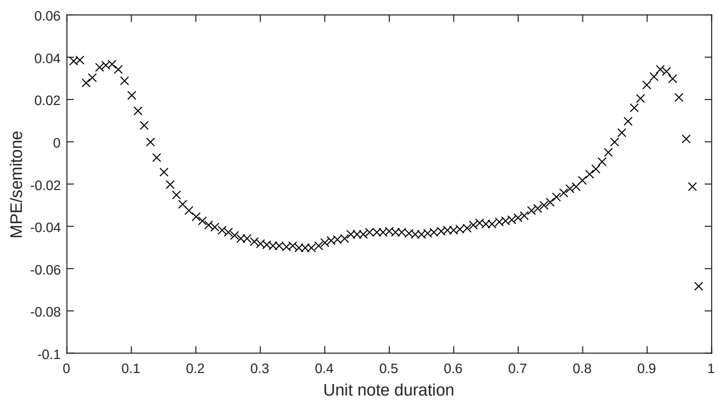

Fig. 1. Plot of mean pitch error in unit duration

For a better understanding of the note trajectories, we extract the first 0.4 and last 0.4 seconds real-time duration (Figure 2 and Figure 3) which is similar trajectories as the unit note. According to the approximate derivatives, the first 0.14 and last 0.14 seconds of note has more pitch variance. The mean note duration is 0.69 seconds where 0.14 seconds duration is about 20% of the unit duration. This result is similar to the proportion of the unit note trajectories. The proportion and direction of beginning and ending are various depending on the individual difference, score pitch, vocal parts. More factors and details will be investigated in the following sections.

The appearance of note trajectories is significantly different between the singers who have a fine music background and the singers who have a few music training. For the good singers, the note trajectories are smooth, two transient parts are quite obvious. For singers has relative less training, their note trajectories tend to uneven and less obvious in the beginning and ending. Figure 4 shows two example of the two bass singers, which have the highest score in

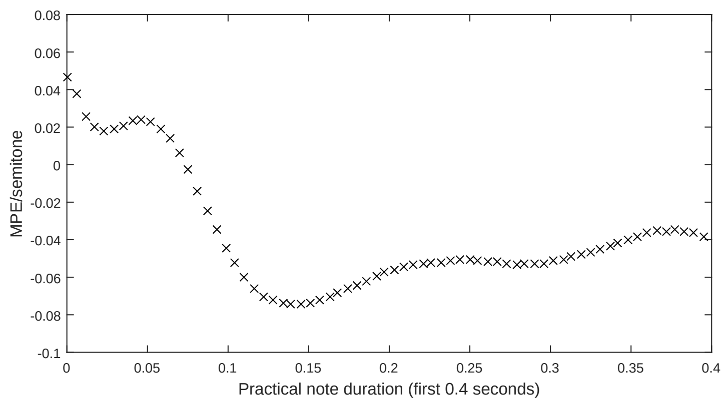

Fig. 2. Practical note duration (first 0.4 seconds)

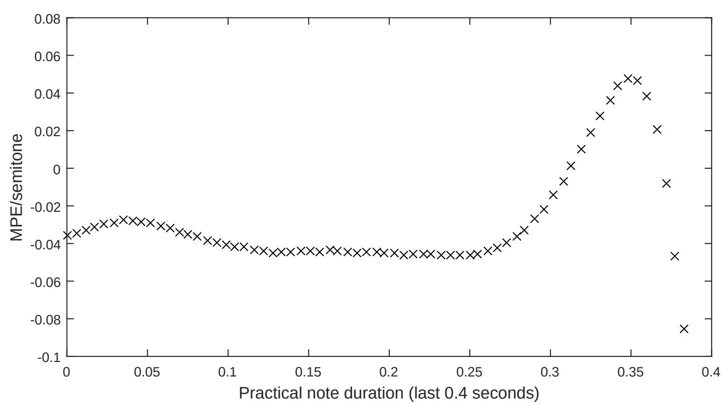

Fig. 3. Practical note duration (last 0.4 seconds)

Golden Smith questionnaire and the lowest one. The zigzag shape and relatively smaller pitch difference inside the note of the less trained singer may due to the unstable pitch and pitch inaccuracy.

In the Figure 2, there is a transient part before the first 0.02 seconds of the real-time duration which may due to the overlap of all the notes, it seems the overall trajectories is the overlap of different types of notes. There are few assumptions of this situation, one is the beginning transient depends on the pitch of the previous note, another is the beginning transient has different tendency according to other factors.

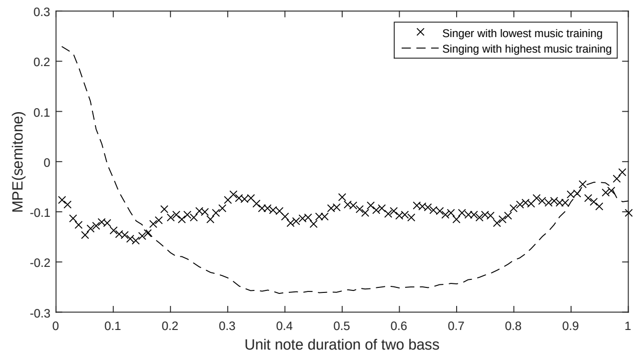

Fig. 4. Unit note duration

4.2 Adjacent pitch

From the previous results, we observed two fluctuations in the beginning and end of the note duration. This fluctuation may because of the adjacent pitch in the score. Then we separate two situations: the previous pitch is lower or higher than current pitch, the next pitch is lower or higher than the current pitch. The Analysis of variance (ANOVA) shows the unit note duration is significantly different between singing before or after a higher or lower pitch ($\mathtt{p} < 0.001$). From the Figure 5 and Figure 6 we found that singers tend to overtone a pitch after singing a low pitch while sing flat after a high pitch. They also prepare the pitch of next note at the end of the previous note. The situation differs according to the individual singers. Most of the singers perform significantly different according to the adjacent notes.

Although the peak might be smaller if the note after a high pitch or has a low pitch adjacent, the slope tendency is still obvious. Some participants have a slope at the beginning while others have a convex in the beginning, there might be more influencing factors rather than the adjacent pitch.

5 Discussion

All the note tends to have a negative transient part at the end, regardless the next pitch is higher or lower. This may due to the human physiological structure. It seems pretty hard to avoid the pitch drift down at end of the note in all the listening conditions, training background and vocal parts. This may indicate that the negative transient at the end does not due to the tone memory or muscle memory of the singers. It is a common phenomenon that prevalent exists.

Besides the music training and adjacent pitches, there are more factors may influence the note pattern. We find that vocal parts and gender also have significant different in note trajectories in terms of transient shape. This may due to the music training of male singers (we observed that bass vocal part has lowest mean pitch error than other three vocal parts in this study). Sopranos have

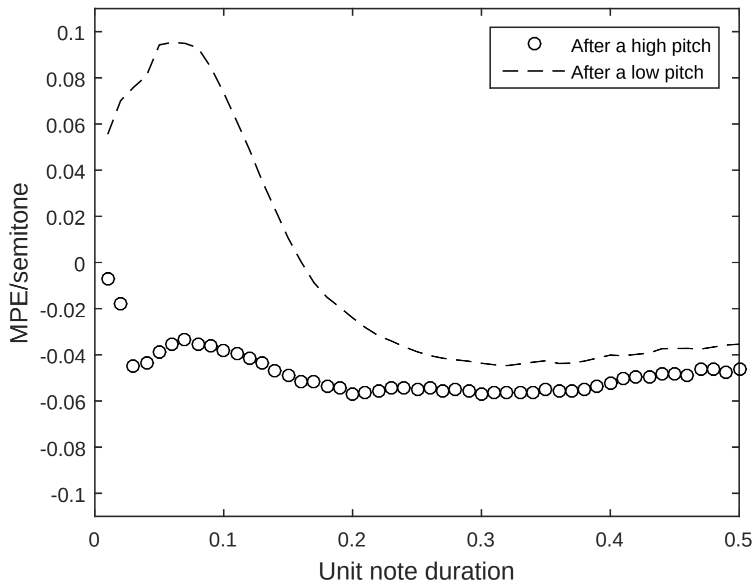

Fig. 5. Unit note duration singing after a lower or higher pitch

highest marks of musical training background in the Golden Smith questionnaire, while the lowest two singers are in the bass part. More investigation of influencing factors will in future works.

Current segmentation is base on the default setting of the software Tony, which separates the notes according to the volume and energy. Different setting and segmentation may influence the results. By randomly selected few samples, the segmentation was been checked. Most of the segmentation seems reasonable, although few notes have a vibration at the end which is hard to tell its tendency, the overlap of thousands of notes shows the negative transient at the end. Only a few of the vague notes were manually segmented which is too small to change the results.

6 Conclusions

In this paper, we represent a study of pitch trajectories inside the single notes. According to the accumulation of over 40 thousands individual notes, we find a general pattern of vocal notes which observes both transient parts in the beginning and end of a note.

The analysis basis on both unit duration and practical duration. The beginning transient is about 15-20%, while the ending transient is about 10-20%. The shape of transient parts are various according to the the individual difference, for example the music training background. The amplitude of the transient parts

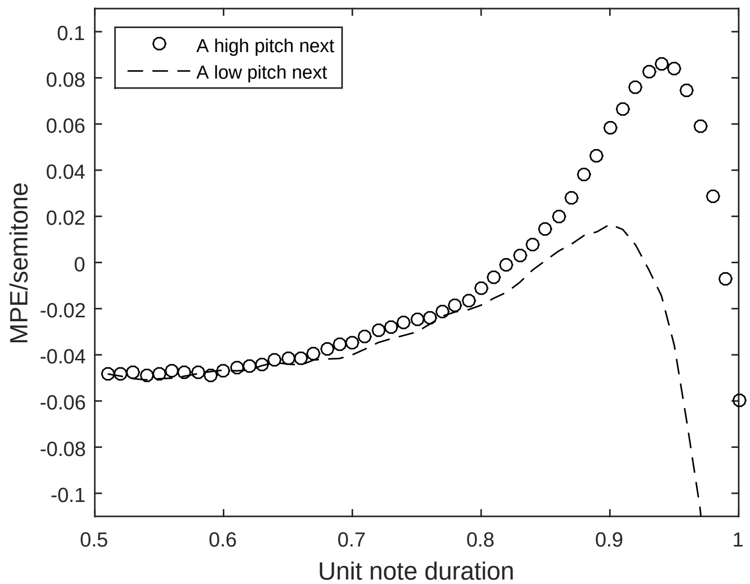

Fig. 6. Unit note duration singing a lower or higher pitch next

were influenced by the adjacent pitch.Participants tend to over tone after a low pitch and arise current pitch at the end of the note if the next pitch is higher.

In conclusion, the main contribution of this paper is the observation, measurement and analysis of the note transient parts by investigating the patterns and influencing factors. Although many further works remain to investigate, we hope that current observation provides a better understanding of singing education and vocal synthesis.

Bibliography

[1] Per-Gunnar Alldahl. *Choral Intonation.* Gehrmans, Stockholm, Sweden, 2008.

[2] Brian J Brandler and Zehra F Peynircioglu. A comparison of the efficacy of individual and collaborative music learning in ensemble rehearsals. *J. Res. Music Educ.*, 63(3):281–297, 2015.

[3] D.E. Brown. *Human Universals.* Temple University Press, Philadelphia, 1991. pp. 1–160.

[4] Nancy A Cooper. Children's singing accuracy as a function of grade level, gender, and individual versus unison singing. *Journal of Research in Music Education*, 43(3):222–231, 1995.

[5] Jiajie Dai and Simon Dixon. Analysis of vocal imitations of pitch trajectories. In *17th Int. Soc. Music Inf. Retr. Conf.*, pages 87–93, 2016.

[6] Jiajie Dai and Simon Dixon. Analysis of interactive intonation in unaccompanied SATB ensembles. In *18th Int. Soc. Music Inf. Retr. Conf.*, pages 599–605, 2017.

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

[8] Joyce Bourne Kennedy and Michael Kennedy. *The Concise Oxford Dictionary of Music.* Oxford University Press, 2004.

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

[10] M. Mauch, C. Cannam, R. Bittner, G. Fazekas, J. Salamon, J. Dai, J. Bello, and S. Dixon. Computer-aided melody note transcription using the Tony software: Accuracy and efficiency. In *Proceedings of the First International Conference on Technologies for Music Notation and Representation*, pages 23–30, May 2015.

[11] Matthias Mauch and Simon Dixon. PYIN: A fundamental frequency estimator using probabilistic threshold distributions. In *IEEE Int. Conf. Acoust., Speech, Signal Process.*, pages 659–663, 2014.

[12] Matthias Mauch, Klaus Frieler, and Simon Dixon. Intonation in unaccompanied singing: Accuracy, drift, and a model of reference pitch memory. *J. Acoust. Soc. Am.*, 136(1):401–411, 2014.

[13] Daniel Müllensiefen, Bruno Gingras, Jason Musil, and Lauren Stewart. The musicality of non-musicians: An index for assessing musical sophistication in the general population. *PLoS ONE*, 9(2):e89642, 2014.

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

[15] John Potter. Introduction: singing at the trun of the century. In John Potter, editor, *The Cambridge Companion to Singing*, pages 1–5. Cambridge University Press, Cambridge, UK, 2000.

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

[17] Johan Sundberg, Filipa MB Lã, and Evangelos Himonides. Intonation and expressivity: A single case study of classical western singing. *Journal of Voice*, 27(3):391–e1, 2013.
[18] Julia Swannell. *The Oxford Modern English Dictionary.* Oxford University Press, USA, 1992. p. 560.
[19] Annie H Takeuchi and Stewart H Hulse. Absolute pitch. *Psychol. Bull.*, 113(2):345, 1993.
[20] Marti Umbert, Jordi Bonada, Masataka Goto, Tomoyasu Nakano, and Johan Sundberg. Expression control in singing voice synthesis: Features, approaches, evaluation, and challenges. *IEEE Signal Processing Magazine*, 32(6):55–73, 2015.
[21] Jean Mary Zarate and Robert J Zatorre. Experience-dependent neural substrates involved in vocal pitch regulation during singing. *Neuroimage*, 40(4):1871–1887, 2008.