

A STUDY OF INTONATION IN THREE-PART SINGING USING THE AUTOMATIC MUSIC PERFORMANCE ANALYSIS AND COMPARISON TOOLKIT (AMPACT)

Johanna Devaney
CNMAT, UC Berkeley/
School of Music
The Ohio State University
j@devaney.ca

Michael Mandel
Audience Inc./
College of Engineering
The Ohio State University
mim@mr-pc.org

Ichiro Fujinaga
CIRMMT
Schulich School of Music
McGill University
ich@music.mcgill.ca

ABSTRACT

This paper introduces the Automatic Music Performance Analysis and Comparison Toolkit (AMPACT), is a MATLAB toolkit for accurately aligning monophonic audio to MIDI scores as well as extracting and analyzing timing-, pitch-, and dynamics-related performance data from the aligned recordings. This paper also presents the results of an analysis performed with AMPACT on an experiment studying intonation in three-part singing. The experiment examines the interval size and drift in four ensembles' performances of a short exercise by Benedetti, which was designed to highlight the conflict between Just Intonation tuning and pitch drift.

1. INTRODUCTION

In the early 20th century, psychologist Carl Seashore and his colleagues at the University of Iowa undertook extensive work in performance analysis of singing, examining dynamics, intonation, and vibrato [22]. Their analyses were based on amplitude and frequency information extracted from recordings with phonophotographic apparatus. These manual methods were extremely labourious and limited the number of recordings that could be accurately analyzed. Recent developments in digital signal processing have allowed for many of these manual processes to be performed computationally.

The MATLAB-based Automatic Performance Analysis and Comparison Toolkit (AMPACT) collects existing tools and introduces a new MIDI-audio alignment algorithm. The alignment algorithm is able to accurately identify onsets and offsets in the difficult cases of the singing voice and instruments with non-percussive onsets and can be trained to work on recordings of a range of voice types and instruments. The analysis portion of the toolkit includes tools for extracting of various performance parameters related to timing, pitch, and dynamics. AMPACT also includes tools for comparing data across multiple performances. The purpose of the toolkit is to facilitate empirical analysis of musical performance for those without extensive technical training.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page.

© 2012 International Society for Music Information Retrieval

This paper also presents an experiment on intonation in three-part singing, which used AMPACT to extract and analyze the intonation data. The experiment uses a short exercise by a music theorist, Benedetti (1530–1590), designed to result in varying amounts of pitch drift when different idealized tunings are applied to it. The exercise was performed numerous times by four different ensembles and the resultant recordings were analyzed in terms of melodic/vertical interval tuning and pitch drift.

2. PREVIOUS WORK

2.1 Automatic Performance Data Extraction

Currently, there are no robust automated methods for estimating note onsets and offsets in the singing voice. Although much work has been done in the area of note onset detection [1], accurate detection of onsets for the singing voice and other instruments without percussive onsets is not a solved problem. Collins used a pitch detector for estimating non-percussive onset detection [3]. He improved on the number of onsets detected within a 100 ms tolerance window over the phase deviation approach described in [1] (58% versus 45%), with comparable false positives (36% versus 37%), Friberg, Schoonderwaldt, and Juslin developed an onset and offset detection algorithm that was evaluated on electric guitar, piano, flute, violin, and saxophone [10]. They reported an onset estimation accuracy of 16 ms and an offset estimation accuracy of 146 ms. Toh, Zhang, and Wang describe a system for automatic onset detection for solo singing voice that accurately predicts 85% of onsets to within 50 ms of the annotated ground truth [23]. These algorithms often require a significant amount of manual correction to obtain sufficient accuracy for performance analysis. Furthermore, offset detection is required for measurements related to duration, intonation, vibrato, and dynamics, but most of these algorithms do not provide it.

2.2 Studies of Intonation in Vocal Ensembles

In the absence of robust automated methods for estimating note onsets and offsets, studies of the singing voice have relied on manual annotation of notes' onsets and offsets. Vurma and Ross studied 13 professional singers' melodic intonation in their performances of short exercises using PRAAT for F_0 analysis [2]. They observed that ascending and descending semitones were smaller than

equal temperament and that ascending and descending fifths were larger than equal temperament [25]. Howard studied two *a cappella* SATB quartets and found that they used non-equal temperament with a tendency towards, though not full compliance with, Just-Intonation [13]. He also argued that in pieces with modulation, that since the ensembles used non-equal temperament, pitch drift is necessary for choirs to stay in tune [14]. Howard used electroglottographs to obtain F_0 estimates in order to avoid the complication of polyphonic F_0 estimation.

3. AMPACT

3.1 Overview

AMPACT¹ automatically analyzes performance data from monophonic or quasi-polyphonic recordings. The algorithms included in the toolkit make use of the information available in the score about what notes are expected in the performance and the order in which they will occur. AMPACT provides estimates of note onsets and offsets for tones with non-percussive onsets (e.g., vocalists) that are more robust than existing blind onset detection or alignment algorithms. The analysis portion of the toolkit allows for the extraction of various performance parameters: inter-onset intervals between notes; tempo information; relative dynamic level between notes; mean frequency for each note and interval sizes in cents; and vibrato rate and depth. The statistical tools allow comparisons of different performances of the same musical material or piece. A schematic of the analysis components of AMPACT is shown in Figure 1.

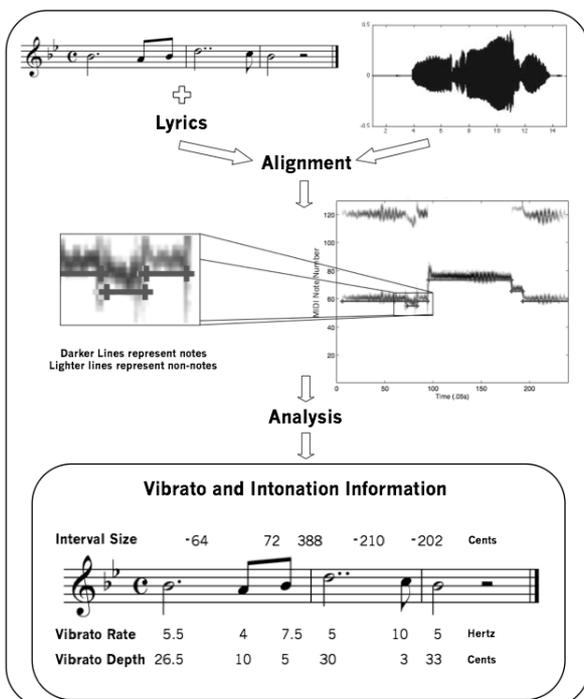


Figure 1. Schematic of the Automatic Performance and Analysis Toolkit (AMPACT).

3.2 MIDI-Audio Alignment

AMPACT uses a MIDI-audio alignment algorithm in order to identify the beginning and ending of all of the notes in a performance. A MIDI version of the score, which is a quantized version of all of the pitch and timing information in the audio, is adjusted so that its timing information corresponds to that of the audio. The algorithm in AMPACT refines the results of an existing Dynamic Time Warping (DTW) approach, described in [18], with a hidden Markov model (HMM). The HMM is trained on the acoustic properties of the melodic line being aligned and, in the case of the singing voice, is guided by the lyrics in the score. The HMM both increases the accuracy of the initial alignment and labels transient and steady-state sections of each note. Identification of the steady-state sections of notes is important because they correspond to the pitched sections.

3.2.1 Hidden Markov Model

This section describes the details of the HMM, with a particular focus on modelling the solo singing voice, namely the observations, states, transition probabilities, and use of a DTW alignment as a prior. The observed variables modelled by the HMM are the square root of periodicity, power, and F_0 estimates provided by the YIN algorithm [5] for each frame. The F_0 estimates from YIN are a somewhat noisy cue, especially for the silence and transient states, but are important because they assist alignment when the note changes under a single vowel.

The three acoustic events are modelled in the HMM: silence, transient, and steady state. In the singing voice, transients occur when a consonant starts or ends a syllable, while vowels produce the steady-state portion of the note. In instruments, the occurrences of transients are influenced by articulation. The transition probabilities were calculated from example recordings of the singing voice. Two versions of the state sequences are implemented. The first allows each state to be visited, shown in Figure 2. The second is modified by the lyrics in the score; transients were only inserted when an unvoiced consonant began or ended a syllable and silences were inserted only at the end of phrases, shown in Figure 3.

The initial DTW alignment is used as a prior to guide the HMM. The use of the DTW alignment obviates the need to encode information about the score in the HMM. By assuming that the DTW alignment is roughly correct, it is not necessary to rely excessively on noisy F_0 estimates in the HMM. This simplifies the design of the HMM and allows the same HMM seed to be used for each note. One issue with this approach is that it cannot adjust the initial alignment by more than one note, so the initial alignment has to be relatively accurate.

¹ Available for download at www.ampact.org

The HMM was implemented in MATLAB with Kevin Murphy’s HMM Toolbox [17] and uses Alain de Cheveigné’s MATLAB implementation of the YIN algorithm [4] as well as Dan Ellis’ MATLAB implementation of DTW MIDI-audio alignment [9]. An evaluation of the alignment algorithm is described in [7].

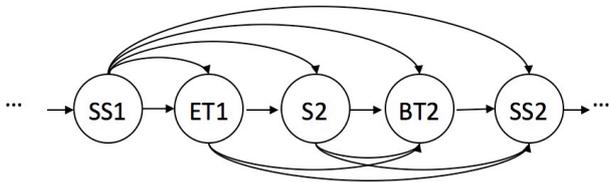


Figure 2. Three-state basic state sequence seed in the HMM: steady state (SS), transient (T), silence (S). The ending transient (ET) and the beginning transient (BT) both have the same observation distribution.

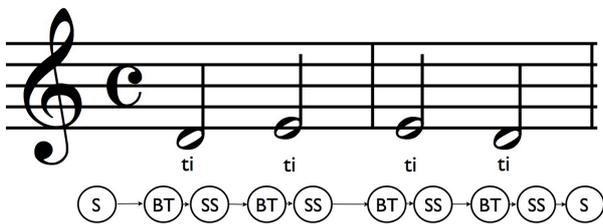


Figure 3. State sequence adapted to sung text.

3.2.2 Performance Data Analysis and Comparison

The alignment algorithm provides information about the note onset and offset times, which AMPACT saves in the MIDI toolbox’s note-matrix format [24] from which a MIDI file can be saved. The onset and offset locations also delineate the starting and ending points for calculating pitch- and dynamics-related parameters of each note. Onset and offset information is also saved in as an Audacity-readable label file [15], which allows for manual correction of any alignment errors AMPACT may make.

The YIN algorithm is used for F_0 estimation. One advantage of YIN is that it allows for specification of minimum and maximum expected F_0 s, which AMPACT sets according to the note information in the aligned score. The maximum F_0 is set to one whole tone above the corresponding note in the score and the minimum F_0 is set to one whole tone below. This is a very useful feature for recordings that are not strictly monophonic, such as recordings from close miking in ensemble performance.

Perceived pitch is calculated using a weighted mean based on the F_0 ’s rate of change [12]. This mean is calculated by assigning a higher weighting to the frames where the F_0 has a lower rate of change and a lower weighting to those with a higher rate of change. The threshold between high and low rates of change is set at 1.41 octaves/second, based on the vibrato rate and depth values reported in [20; 21]. Vibrato is calculated by finding the

dominant frequency of the FFT of the pitch contour. Dynamics are calculated using the implementation in Genesis Acoustics Loudness Toolbox for MATLAB [11] of Glasberg and Moore’s model for estimating loudness in time-varying sounds [16]. AMPACT also includes tools for statistical comparison of performances through a wrapper for various t-test, ANOVA, and linear regression functions in MATLAB.

4. INTONATION EXPERIMENT

AMPACT was used to extract and analyze intonation data in an experiment on four three-part vocal ensembles. The ensembles’ performances were analyzed with regard to melodic whole-tone tunings, a range of vertical interval tunings, and overall pitch drift. A pre-release version of AMPACT was used by the authors in larger-scale experiment on solo singing in [6].

4.1 Method

4.1.1 Experimental Material

The experimental material is a three-part chord progression written by Benedetti that was designed to show that singers would not tune Justly with the current sustained note since strict adherence to Just Intonation would result in a significant pitch drift that is not observable in performances of the progression [19]. The progression is built from a seed two-measure progression that is repeated four times. If the singers were to tune in Just Intonation to the sustained note, rather than the bass note, the ensemble would drift up a syntonic comma (21.5 cents) by the end of each seed, resulting in a total upwards drift of 86 cents by the end of the four repetitions. In contrast, if the singers were to tune to the bass in each vertical sonority, with D, A, or G in the bass, there should be no drift. The calculations for both tuning scenarios are shown in Figure 4.

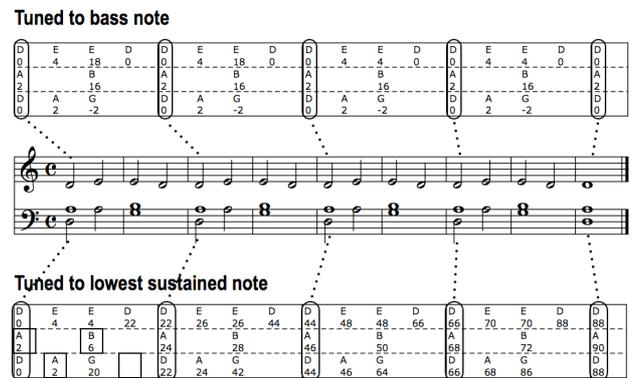


Figure 4. Theoretical tuning for Benedetti progression used as experimental material. The numbers in the tables at the top and bottom of the figure indicate the tuning in relation to the starting pitch in the bass.

4.1.2 Participants

Four three-part ensembles participated in this experiment. Ensemble 1 served as a pilot study with semi-professional alto, tenor, and bass singers who performed without a conductor. The ensemble had an average age of 26 years ($SD = 3.6$), with an average of 6.5 years of private voice lessons ($SD = 4.5$) and 6.5 years of regular practice ($SD = 2.5$). Ensembles 2, 3, and 4 consisted of professional singers who regularly sang together with the conductor used in the experiment. These ensembles had an average age of 42 years ($SD = 9$), an average of 7.75 years of private voice lessons ($SD = 0.5$) and 24 years of regular practice ($SD = 10$). The singers in both ensembles were experienced in singing *a cappella* Renaissance music and were asked to sing with their normal tuning.

Ensembles 2 and 4 consisted of alto, tenor, and bass singers while Ensemble 3 consisted of soprano, alto, and tenor. Ensembles 1 and 2 were recorded in a 4.85m x 4.50m x 3.30m lab with low-noise, minimal reflections, and short reverberation time. The singers were miked with cardioid headband mics (DPA 4088-F). The microphones were run through an RME Micstasy 8-channel microphone preamplifier and an RME Madi Bridge into a Mac Pro computer for multi-track recording. Ensembles 3 and 4 were recorded on the altar of St. Mathias' Church, a church in Montreal dating from 1912 with wooden floors, limestone walls, and seating for 350 people. As with the lab environment, the singers were miked with cardioid headband mics, although a portable Zaxcom Deva 16 digital recorder was used for the rest of the recording setup.

4.2 Results

4.2.1 Interval Size

The mean and standard deviation of the interval sizes for the melodic whole tones are shown in Table 1. The majority of the means were within one standard deviation of the equal tempered 200 cent tuning. The main exception to this was Ensemble 1 where the whole tone tended to be smaller. In particular, the middle voice whole tones, which were closer to the 182 cent Minor Just Intonation whole tone. Just over half of the singers' whole tones (12/20) were within one standard deviation of the Pythagorean/Major Just Intonation (204 cents) whole tone.

Vertical intervals were calculated for each half-measure between all of the voices: lowest voice to middle voice, lowest voice to upper voice, and middle voice to upper voice. The onset and offset times for the vertical intervals were determined by the upper voice. Overall, there were 51 vertical intervals in each rendition: 4 Minor Thirds (m3), 8 Major Thirds (M3), 9 Perfect Fourths (P4), 17 Perfect Fifths (P5), 4 Major Sixths (M6), and 9 Perfect Octaves (P8). The means and standard deviations for each type of vertical interval across all of the singers

in each ensemble are shown in Table 2. There was a wide range in the mean values for both the vertical m3 and M3, specifically 300–322 cents for the m3 and 375–413 cents for the M3. When the standard deviations are taken into account, the m3 encompassed the Pythagorean (294 cent), equal tempered (300 cents), and Just Intonation (316 cents) tunings. Likewise, the M3 range encompassed the Just Intonation (386 cents), equal tempered (400 cents), and Pythagorean (408 cents) tunings. The range of the means for the M6 encompassed only the equal tempered tuning (900 cents) since the means were all larger than the Just Intonation tuning (884 cents) and marginally smaller than the Pythagorean (905 cents). The tunings for the P4 (498 cents), P5 (702 cents), and P8 (1200 cents) are common to both the Pythagorean and Just Intonation systems and are close to the values for equal temperament (500, 700, and 1200 cents, respectively); the ranges for these intervals encompassed all of these tunings.

		Voices				
		Top		Middle		Bottom
Ensemble		Up	Down	Up	Down	Down
1	Mean	199	195	185	183	191
	SD	6	4	6	10	7
	N	12	12	12	12	12
2	Mean	192	191	207	210	189
	SD	6	16	12	13	6
	N	16	16	16	16	16
3	Mean	199	199	199	196	198
	SD	9	10	8	8	11
	N	16	16	16	16	16
4	Mean	189	194	196	196	195
	SD	13	8	10	13	13
	N	20	20	20	20	20

Table 1. Mean, standard deviation, and number of instances of the ascending and descending melodic whole tone sizes for all ensembles, broken down by voice.

		Vertical Interval Types					
		m3	M3	P4	P5	M6	P8
Ensemble							
1	Mean	322	376	509	701	893	1201
	SD	7	9	10	6	6	7
	N	12	24	27	51	12	27
2	Mean	300	413	497	705	903	1206
	SD	12	11	17	14	15	12
	N	16	32	36	68	16	36
3	Mean	307	397	507	704	904	1209
	SD	8	11	12	11	11	9
	N	16	32	36	68	16	36
4	Mean	301	406	493	702	896	1202
	SD	14	15	13	12	10	12
	N	20	40	45	85	20	45

Table 2. Mean, standard deviation, and number of instances of the vertical interval sizes between the three voices across all renditions by all of the ensembles.

An ANOVA analysis for each ensemble was run on the melodic interval data with whole tone size as the dependant variable and direction and singer identity as independent variables. In Ensemble 1, there was no significant effect for direction, though the middle singer's whole tones were significantly smaller than the other two singers, $F(2, 56) = 24.59$, $p < 0.001$. Ensemble 2 was similar, with no effect for direction and a significant effect for the middle singer, though in this case the middle singer's whole tones were significantly larger than the other two singers, $F(2, 75) = 24.52$, $p < 0.001$. There were no significant effects for direction or singer identity in Ensembles 3 and 4. A separate ANOVA was run with direction and group identity. There was no overall effect for direction, but Ensemble 1's whole tones were significantly smaller on average than Ensembles 2 and 3, $F(3, 311) = 6.96$, $p < 0.001$.

Separate ANOVAs were also run on each vertical interval to test for group effects. Ensemble 1's m3 intervals were significantly larger on average than the other three ensembles, $F(3, 59) = 11.93$, $p < 0.001$, so much so that their mean overshot the 316 cent Just Intonation value. In contrast, Ensemble 1's M3 intervals were significantly smaller on average than the other ensembles', $F(3, 127) = 50.31$, $p < 0.001$ and were so small that they undershot the 386 cent Just Intonation value. For the P4, Ensembles 1 and 3 were significantly larger than Ensembles 2 and 4, $F(3, 143) = 11.75$, $p < 0.001$. There were no significant effects between the ensembles for the P5, M6, or P8.

4.2.2 Pitch Drift

In order to assess whether the ensembles drifted in the manner predicted by Benedetti, the perceived pitch estimates in cents were calculated for each note in each rendition in relation to the rendition's opening D in the bass. The table at the top of Figure 5 shows the means and standard deviations for each note across all of the ensembles. Overall there was a slight drift upwards of 8 cents in the lower voice, 10 cents in the middle voice, and 13 cents in the upper voice. This drift is much smaller than the one predicted by the calculations in the lower chart in Figure 5, suggesting that the singers were tuning to the bass note rather than the lowest sustained note.

Figure 5 also shows the drift in the bass voice for each ensemble through plots of the perceived pitch (relative to the starting note) of the D at the start of each seed progression. Ensemble 1 was the most consistent with itself across performances, exhibiting only a small amount of drift from the starting pitch. Ensembles 2 and 3 both tended to drift upwards with Ensemble 3 showing a greater amount of variability in the amount of drift. Ensemble 4 had little drift overall but showed a large amount of variation within each performance.

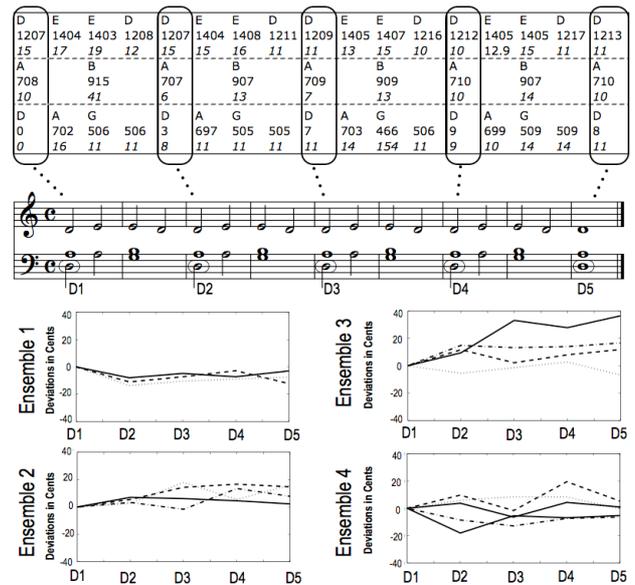


Figure 5. Summary of the amount of drift in each ensemble's renditions of the Benedetti chord progression. The lines in the each plot link the perceived pitch estimates for the notes D1-D5 in each rendition.

4.3 Discussion

Overall the singers tended towards equal temperament. The vast majority of the means of the melodic and vertical intervals were within one standard deviation of equal temperament. The melodic intervals that were not within one standard deviation of equal temperament were much smaller, falling instead within one standard deviation of the 182 cent minor Just Intonation semitone. Likewise, most of the outlying vertical intervals fell within one standard deviation of non-equal temperament idealized tunings (either Just Intonation or Pythagorean): Ensemble 1's m3 mean was within one standard deviation of the 316 cent Just Intonation tuning; Ensemble 1's M3 mean was within one standard deviation of the 386 cent Just Intonation tuning; Ensemble 2's M3 mean was within one standard deviation of the 408 cent M3. The ANOVA analysis revealed some significant effects for singer and group identity for some of the ensembles. The lack of a significant effect for direction in the ANOVA analysis of the whole tone tuning mirrors our earlier findings for professional solo singers [8].

The ensembles did drift up slightly on average, but not to the extent they would have if the singers were tuning in Just Intonation to the lowest sustained note. This is not surprising, as such a rapid drift, 88 cents over eight measures, is highly unlikely since it implies that the singers were not retaining their starting pitch as a reference only a few tens of seconds after it was sung.

5. CONCLUSIONS

This paper presented AMPACT, a MATLAB toolkit for automatically extracting, analyzing, and comparing per-

formance data from monophonic recordings for which a score is available. The alignment algorithm in AMPACT works well on sounds without a clearly defined onset, making it useful for the singing voice and instruments with non-percussive onsets. This paper also demonstrates the use of AMPACT in extracting and analyzing the data for an experiment on vocal intonation. The experiment with four three-part ensembles found that the singers tended toward equal temperament and did not exhibit a large amount of drift in an exercise by Renaissance theorist Benedetti. The detailed analysis of singing intonation in this study was facilitated by the automated nature of AMPACT, not only in terms of time savings but also in consistency of data extraction.

6. ACKNOWLEDGMENTS

The authors would like to thank Dan Ellis for his assistance in developing the alignment algorithm in AMPACT. We would also like to thank Jonathan Wild and Peter Schubert for their contributions to the experiment. We would also like to acknowledge the contributions of the personal at the Centre for Research in Music Media and Technology (CIRMMT) and Center for New Music and Audio Technologies (CNMAT) to various aspects of this project, particularly David Wessel. This work was supported by funding from the Fonds de recherche sur la société et la culture (FQRSC) and Social Sciences and Humanities Research Council of Canada (SSHRC).

7. REFERENCES

- [1] Bello, J. P., L. Daudet, S. Abdallah, C. Duxbury, M. Davies, and M. Sandler. 2005. A tutorial on onset detection in music signals. *TASLP* 13 (5): 1035–47.
- [2] Boersma, P. 2001. PRAAT, a system for doing phonetics by computer. *Glott International* 5 (9/10): 341–5.
- [3] Collins, N. 2005. Using a pitch detector for onset detection. In *Proceedings of ISMIR*, 100–6.
- [4] de Cheveigné, A. 2002. *YIN MATLAB implementation*. <http://audition.ens.fr/adc/sw/yin.zip>.
- [5] de Cheveigné, A., and H. Kawahara. 2002. YIN, a fundamental frequency estimator for speech and music. *JASA* 111 (4): 1917–30.
- [6] Devaney, J., M. Mandel, I. D. P. W. Ellis, and I. Fujinaga. 2011. Automatically extracting performance data from recordings of trained singers. *Psychomusicology* 21 (1–2): 108–36.
- [7] Devaney, J., M. I. Mandel, and D. P. W. Ellis. 2009. Improving MIDI-audio alignment with acoustic features. In *Proceedings of WASPAA*, 45–8.
- [8] Devaney, J., J. Wild, and I. Fujinaga. 2011. Intonation in solo vocal performance: A study of semitone and whole tone tuning in undergraduate and professional sopranos. In *Proceedings of the International Symposium on Performance Science*, 219–24.
- [9] Ellis, D. P. W. 2008. *Aligning MIDI scores to music audio*. <http://www.ee.columbia.edu/~dpwe/resources/matlab/alignmidiwav/>.
- [10] Friberg, A., E. Schoonderwaldt, and P. N. Juslin. 2007. CUEx: An algorithm for extracting expressive tone variables from audio recordings. *Acta Acustica united with Acustica* 93: 411–20.
- [11] *Genesis Acoustics Loudness Toolbox for Matlab*. 2010. <http://www.genesis-acoustics.com/>.
- [12] Gockel, H., B. C. J. Moore, and R. P. Carlyon. 2001. Influence of rate of change of frequency on the overall pitch of frequency-modulated tones. *JASA* 109 (2): 701–12.
- [13] Howard, D. M. 2007. Equal or non-equal temperament in a cappella SATB singing. *Logopedics Phoniatrics Vocology* 32: 87–94.
- [14] Howard, D. M. 2007. Intonation drift in a capella soprano, alto, tenor, bass quartet singing with key modulation. *Journal of Voice* 21 (3): 300–15.
- [15] Mazzoni, D., and R. Dannenberg. 2000. *Audacity*. <http://audacity.sourceforge.net/>.
- [16] Moore, B. C. J., and B. R. Glasberg. 2010. The role of temporal fine structure in harmonic segregation through mistuning. *JASA* 127 (1): 5–8.
- [17] Murphy, K. 1998. *Hidden Markov Model (HMM) Toolbox for Matlab*. <http://www.cs.ubc.ca/~murphyk/Software/HMM/hmm.html>.
- [18] Orio, N., and D. Schwarz. 2001. Alignment of monophonic and polyphonic music to a score. In *Proceedings of ICMC*, 155–8.
- [19] Palisca, C. V. 1994. *Studies in the History of Italian Music and Music Theory*. Oxford, UK: Oxford University Press.
- [20] Prame, E. 1994. Measurements of the vibrato rate of ten singers. *JASA* 96 (4): 1979–84.
- [21] Prame, E. 1997. Vibrato extent and intonation in professional western lyric singing. *JASA* 102 (1): 616–21.
- [22] Seashore, C. 1938. *Psychology of Music*. Iowa City, IA: University of Iowa Press. Original edition, New York, NY: Dover Publications.
- [23] Toh, C. C., B. Zhang, and Y. Wang. 2008. Multiple-feature fusion based on onset detection for solo singing voice. In *Proceedings of ISMIR*, 515–20.
- [24] Toivainen, P., and T. Eerola. 2004. *MIDI Toolbox: MATLAB Tools for Music Research*. <http://www.jyu.fi/musica/miditoolbox/>
- [25] Vurma, A., and J. Ross. 2006. Production and perception of musical intervals. *Music Perception* 23 (4): 331–44.