The Zurich Corpus of Vowel and Voice Quality, Version 1.0

Dieter Maurer[1], *Christian d'Heureuse*[1], *Heidy Suter*[1], *Volker Dellwo*[2], *Daniel Friedrichs*[3], *Thayabaran Kathiresan*[2]

[1]Institute for the Performing Arts and Film, Zurich University of the Arts, Switzerland
[2]Department of Computational Linguistics, University of Zurich, Switzerland
[3]Department of Speech, Hearing and Phonetic Sciences, University College London, United Kingdom

dieter.maurer@zhdk.ch, chdh@inventec.ch, heidy.suter@zhdk.ch, volker.dellwo@uzh.ch, daniel.friedrichs@ucl.ac.uk, thayabaran.kathiresan@uzh.ch

Abstract

Existing databases of isolated vowel sounds or vowel sounds embedded in consonantal context generally document only limited variation of basic production parameters. Thus, concerning the possible variation range of vowel and voice quality-related sound characteristics, there is a lack of broad phenomenological and descriptive references that allow for a comprehensive understanding of vowel acoustics and for an evaluation of the extent to which corresponding existing approaches and models can be generalised. In order to contribute to the building up of such references, a novel database of vowel sounds that exceeds any existing collection by size and diversity of vocalic characteristics is presented here, comprised of c. 34 600 utterances of 70 speakers (46 non-professional speakers, children, women and men, and 24 professional actors/actresses and singers of straight theatre, contemporary singing, and European classical singing). The database focuses on sounds of the long Standard German vowels /i–y–e–ø–ɛ–a–o–u/ produced with varying basic production parameters such as phonation type, vocal effort, fundamental frequency, vowel context and speaking or singing style. In addition, a read text and, for professionals, songs are also included. The database is accessible for scientific use, and further extensions are in progress.

Index Terms: vowel acoustics, vowel recognition, voice quality, speech perception, speech databases

1. Introduction

Besides the great many databases of continuous speech, numerous smaller samples or larger databases of vowel sounds produced in isolation (V context), in minimal pairs (e.g., hVd) or in nonsense syllables are also reported in the literature, and some of them are accessible. They generally serve as an empirical basis to address specific issues regarding acoustic and perceptual characteristics of vowel and voice quality. Among the many topics addressed, major questions concern (i) the assessment of language-specific average formant patterns for relaxed speech or for sounds in citation-form words, respectively [e.g., 1, 2] including testing the maintenance or alteration of vowel quality in resynthesis [e.g., 3, 4], (ii) the comparison of estimated formant patterns versus whole spectral envelopes [e.g., 4, 5] and static versus dynamic acoustic characteristics as representing vowel quality [e.g., 6–8], (iii) normalisation procedures for age- and gender-related acoustic differences (see [9] for an overview) including the investigation of the perceptual effect of fundamental frequency (f_o) [e.g., 10], (iv) dialect variation [e.g., 11–13], (v) phonation type specific acoustic characteristics [e.g., 14–16], (vi) the effect of vocal effort on spectral properties of vowels [e.g., 17, 18], (vii) the effect of vowel duration [e.g., 19] and consonantal context [e.g., 20] on vowel recognition, (viii) aspects of voice source characteristics [e.g., 21], and (ix) automatic vowel and word or speaker or voice quality classification and recognition procedures [e.g., 22–25]. Besides, some samples also document vowel sounds sung by professionally trained singers (for an overview see [26]; see also [27, 28]).

However, in general, existing samples or databases either present sounds produced by a given speaker with medium vocal effort at particular f_o, or they compare sounds of a given speaker related to only two different production parameters, e.g., voiced and whisper phonation, or V and CVC context, or voiced with varying vocal effort, or voiced with varying f_o in singing, etc. (for corresponding references see above). To the best of our knowledge, no database exists that includes an extensive and combined variation of basic production parameters such as phonation type, vocal effort, f_o, and vowel context for the sounds of each single documented speaker. Therefore, we do not have phenomenological and descriptive references at our disposal that allow for a comprehensive understanding of the acoustics and perception of vowel and voice quality and for an evaluation of the extent to which corresponding existing approaches and models can be generalised. Yet, such a comprehensive understanding is needed because the many studies on the matter have shown a strong effect of production parameter variation on acoustic properties and sometimes also on perceptual characteristics, and they have pointed towards arising methodological problems.

Against this background, we have built up a large database of vowel sounds that includes an extensive and systematic variation of basic production parameters. In addition, we have further included the comparison of untrained non-professional speakers (hereafter: non-professionals) and trained and professionally active speakers and singers (hereafter: professionals) the latter representing three different artistic production styles, straight theatre (ST), contemporary singing (CS, for substyles see below), and European classical singing (EC).

The database appertains a double structure of investigation and documentation. The main body of extensive investigation and documentation focusses on recordings of utterances in Standard German produced by 40 non-professionals and professionals, on the sound production of the long vowels /i–y–e–ø–ɛ–a–o–u/ and varying phonation type, vocal effort, f_o,

vowel context, and speaking or singing style. A read text and, for the professionals, one or several songs are also included. This main body is comprised of c. 33 800 recordings. – The side body consists of reference recordings of the same set of vowel sounds (see above), produced by 30 non-professionals, with no production parameter variations except f_0 variations within an everyday speaking range. This side body is comprised of 830 recordings.

Below, the method applied and the design of the database are reported in general terms. For further details and access to the database see [29].

2. Method

2.1. Speakers and utterances

Speakers of extensive investigation (main body): As shown in Table 1 (see speaker groups A to D), 16 non-professionals (8 children, aged 7 to 10, and 8 adults, aged 23 to 40, gender balanced) and 24 professionals (adults, aged 25 to 56, gender balanced) with no report of hearing impairment were investigated concerning utterances with extensive varying basic production parameters. The professional group is comprised of 8 ST actresses/actors, 8 CS singers (including the substyles contemporary musical theatre, pop, and jazz), and 8 EC singers (2 sopranos, 2 mezzo-sopranos, 2 tenors, and 2 baritones). – Non-professionals were selected according to two criteria: a minimal vocal range for vowel production of 24 semitones (2 octaves) for adults and 19 semitones (c. 1.5 octaves) for children, with vowels recognisable over a range of 15 semitones in minimum for both adults and children (for details see the handbook in [29]). Professionals were selected according to their professional status, their praxis of performing in Standard German, their willingness to participate in a scientific investigation and their geographic reachability. The professional status was assigned according to Bunch and Chapman [30], with ranking levels 2 or 3 of this taxonomy. – The speaker selection was made by the first and the third author, both trained singers. – All speakers are native speakers of German, with origins in Germany, Austria or the northern part of Switzerland, with the exception of 4 professionals (all singers), not being native speakers of German, but professionally performing on stage in Standard German. – All adult speakers were remunerated with a participation fee. The children obtained a small gift.

Table 1: *Speaker subgroups and speaker numbers.*

Speaker groups	Children f	Children m	Adults f	Adults m	Speakers total
Main body					
A : Non-professionals	4	4	4	4	16
B : ST actresses/actors	–	–	4	4	8
C : CS singers	–	–	4	4	8
D : EC singers	–	–	4	4	8
Side body					
E : Non-professionals	5	5	10	10	30
Total entire database					**70**

Utterances of extensive investigation (main body): As shown in Table 2, the speakers of extensive investigation produced sustained sounds of the 8 long Standard German vowels /i–y–e–ø–ɛ–a–o–u/ with varying basic production parameters for phonation (voiced, breathy, creaky, whisper), vocal effort (medium, low, high, shouted), vowel context (V and sVsV), and f_0 (monotonous f_0 levels according to C-major scale, covering the most of the speaker's vocal range; exceptions were shouted sounds on freely-chosen pitches). All utterances were made by the speakers as non-professional (non-style) productions, that is favouring the intelligibility of vowel quality over sound timbre. Consequently, and most importantly, the professionals had to attempt to partially or fully abandon their style training. – In addition to the non-style utterances, the professionals were also asked to produce the set of voiced sounds with corresponding production parameter variation in their own respective style of singing or speaking for a range of f_0 that reflects their artistic style. Thus, the vowel production of the professionals was investigated with regard to both their attempt at producing clearly recognisable vowel sounds as well as a performance in their respective professional style. – The production of vowel sounds in sVsV context was limited to voiced sounds on a higher f_0 range ($\geq$ 523 Hz for children and women, $\geq$ 330 Hz for tenors and high male voices, $\geq$ 262 Hz for baritones and middle male voices) and to shouted and whispered sounds, since consonantal context was investigated only in terms of crosschecking its role for vowel recognition concerning three kinds of possibly critical vowel sound production: high pitch range, very high vocal effort, and whispering. – For sound duration see below.

The non-professionals also read a reference text ("Nordwind und Sonne") and sang a song in German. The professionals read the same text in non-style as well as in style mode and sang a song in German in their respective singing style. For some speakers of CS and EC styles, additional songs in Italian and English were also recorded.

Table 2: *Production parameter configurations.*

Speaker groups	Phonation	Vocal effort	Sound context	f_0 variation
Main body	voiced	medium	V	scale
A-D	voiced	low	V	scale
	voiced	high	V	scale
	voiced	medium	sVsV	upper scale
	voiced	shouted	V	-
	voiced	shouted	sVsV	-
	breathy	-	V	-
	creaky	-	V	-
	whisper	-	V	-
	whisper	-	sVsV	-
Side body	voiced	medium	V	reference f_0
E				

Reference speakers (side body): Vowel production with very limited f_0 variation of 30 native Swiss German non-professionals (10 children, aged 7–9, and 20 adults, aged 18–52, gender balanced) were also investigated (see Table 1, group E). Speakers were selected according to their native Swiss German dialect in the Northern part of Switzerland, their command of speaking Standard German (primary language in school in Switzerland) and their ability to produce vowel sounds on a specific pitch over an f_0 range of 15 semitones in minimum. The selection was made by the first and the third author. All speakers participated voluntarily with no remuneration.

Reference utterances (side body): The reference speakers produced sustained sounds of the 8 long Standard German vowels /i–y–e–ø–ɛ–a–o–u/ in isolation (V context) with medium vocal effort on monotonous f_0 levels of 220–262–440–523 Hz for children, 220–262–440 Hz for women and 131–220–262 Hz

for men. f_o variation was included in order to cover the f_o contour of real speech prosody and to allow for a compari-son of sounds on different and similar f_o for the different age and gender subgroups. In addition, the speakers read the reference text (see above). – This sample of reference speakers and utterances, limited to a very restricted geographical region, was collected in order to provide evidence that all speakers of extensive investigation (with less regional restriction of speaker origin) show comparable vowel pronunciation when compared to reference utterances (with narrow regional restriction of speaker origin) both in terms of acoustic characteristics and vowel recognition, given corresponding f_o and vocal effort levels.

2.2. Recordings

Permissions: All speakers were informed in detail about the aim and procedure of investigation and gave a written consent to publish their vocal recordings for all scientific purposes, provided that speaker identification is anonymised. For children, written consent was given by a parent.

Recording setting: Utterances were made in a quiet room in standing position and were digitally recorded (44.1 kHz sampling frequency, 24 bits amplitude resolution, mono) using a cardioid condenser microphone (Sennheiser MKH 40P 48) with a pop screen, mounted on a microphone stand. Speaker–microphone distance was 30 cm. The microphone was connected to a PC via an audio interface (Fireface UCX). Recorded sounds were stored in WAV format.

Calibration of sound levels: Before a sequence of recordings related to specific production parameters, the speaker produced several test vowel sounds on different f_o levels in order to set the microphone input gain to a suitable level for the vocal effort investigated. For the read text and the aria or song, the gain was adjusted in the same manner. In order to subsequently determine the actual sound pressure level, for each recording session, a 1 kHz sine wave was recorded with a reference gain using a sound level calibrator (Brüel & Kiær 4230).

Recording procedure: Utterances were recorded according to a specific configuration of production parameters (see Table 2), separating non-style and style productions. – For sounds in V context, except for shouting, the speakers were asked to monotonously sustain a sound on a given f_o level for more than 1 sec if possible. For sounds in sVsV context, the speakers were asked to monotonously sustain the first or the second vowel in the non-word for more than 0.5 sec if possible. However, the actual sound duration varies strongly among speakers and specific configurations of production parameters. But as a rule, a minimum steady-state vowel nucleus (excluding on-/offsets) of 0.5 sec for sounds in V context and of 0.3 sec for sounds in sVsV context is provided for the sounds published (for exceptions see the handbook in [29]). – Two investigators with extensive singing training and phonetic expertise (first and third author) conducted and supervised the recordings. High attention was paid to not to overstrain vocal performances of the participants and to remain within the range of a healthy voice production even in cases of investigating vocal range limits. – If the speaker or the investigator judged that another utterance could improve the sound or vowel quality, the recording was repeated one more or several more times.

f_o scale: For each speaker, a comfortable "middle" f_o level on the C-major scale was determined from which vocalisations were then produced up and down the C-major scale. If the speaker was familiar with the musical scale, this "middle" f_o level was played back by an electronic piano sound, and the speaker subsequently varied f_o autonomously. If not, each f_o level next on the scale was played back by an electronic piano sound or was vocally presented by the investigator.

Corrections: If possible, after performing a listening test of the vowel qualities of the recorded sounds (see below), some speaker recordings were repeated for sounds with low recognition rate or short duration in order to crosscheck possible improvements. However, corrections were limited to non-style productions only, and they were not feasible for all speakers due to professional engagement and geographic availability.

Recording period: The recordings were made in the time period from 2013 to 2018.

Editing: Single sound files were extracted from the recorded sound series using a semi-automatic tool (proprietary development). The cuts were made so as to include full on- and offset of the sounds and to approximately centre the sound in a single sound file, above all for cases of pronounced asymmetries of on- and offsets. Each single sound file was then labelled with a database reference number and relevant sound and speaker information in anonymous form.

2.3. Acoustic analysis

Analysis: For utterances in V context, the analysis was conducted on the middle 0.3 sec of each isolated vowel sound for a frequency range of 0–5.5 kHz on f_o contour, average f_o frequency, average spectrum, spectrogram, average formant patterns (frequencies, bandwidths, levels), and formant tracks. In addition, the average spectrum was also calculated for a frequency range of 0–11 kHz. – Concerning formant pattern estimation, LPC analysis (Burg algorithm, window length=25 ms, time steps=5 ms, pre-emphasis=50 Hz) was conducted in parallel for three parameter settings according to three commonly used age- and gender-related standards of 12 (standard for men), 10 (standard for women) and 8 (standard for children) poles for the frequency range of 0–5.5 kHz. – The same analysis was conducted on sVsV sounds for the middle 0.3 sec of the first or the second vowel sound, depending on their duration (for details of automatic procedure see [29]). – The read texts and the songs/arias were analysed for f_o contour, spectrogram (0–5.5 kHz) and long-term average spectrum (LTAS; 0–5.5 and 0–11 kHz). – The acoustic analysis was conducted with a script using the Praat functionalities [31].

Graphic representations, numerical indications: For vowel sounds, graphic representation includes the display of the entire sound wave, the sound nucleus analysed, the f_o contour, the spectrum, the spectrogram and the formant tracks. In addition, three LPC filter curves (for the three parameter settings mentioned) of the middle window of the sound nucleus was overlaid to the spectrum in order to illustrate the correspondence between spectral peaks and calculated formants. – Numerical average values of f_o and formant patterns were added to the sound information. – For texts and songs/arias, graphic representation included the display of the sound wave, the f_o contour, the spectrogram and the LTAS.

Crosschecks: Sounds were acoustically crosschecked and sounds with background noise were removed. Graphic representtations and numerical values were visually crosschecked for accuracy of cuts, assignment of 0.3 sec vowel nucleus and calculated average f_o. In cases of f_o calculation errors, calculated f_o levels were manually corrected by ear.

2.4. Listening test

Listeners: Five professionally trained speakers (singers or actors or voice teachers) listened to all vowel sounds to assign a perceived vowel quality. Each listener passed a pure-tone hearing screening (25 dB at octave frequencies from 0.5–4 kHz, using a Beltone 110 audiometer) in order to exclude hearing impairment.

Listening test: Testing vowel recognition was organised into speaker-specific subtests (blocked-speaker condition), further separating non-style and style utterances. The sounds were presented in random order. The listeners performed the listening tests remotely online over the entire recording period, using a personal computer and headphones (Beyerdynamic DT 770 Pro). – Before each subtest, an extract of 50 sounds of this subtest (or, for smaller subtests, all sounds) were played in random order to get familiarised with the speaker's phonation, articulation and production parameter variation. Subsequently, the actual test was performed: listeners were asked to listen to each sound and to assign the perceived vowel quality to either one specific Standard German vowel (/i–y–e–ø–ɛ–a–ɔ–o–u/), to a vowel boundary region of two vowels maximum, to "no vowel", or to a free comment. If a sound was difficult to identify they could listen to it repeatedly. – The vowel /ɔ/ was included in the listening test because the perceptual distance /a–o/ is very large, not representing adjacent vowel qualities. – The assignment of the vowel /a/ included all variants in the region of /ɑ–a/ because the production of this vowel varies strongly among German speakers.

2.5. Sounds recorded and sounds selected for publication

For the production parameters documented in this first version of the corpus, in total, c. 57 500 recordings were made for all 70 speakers. As mentioned, for a specific configuration of production parameters, in many cases, two or multiple recordings were made to obtain the best vowel or sound quality. For the publication of the open accessible database, a subset of the recorded sounds was selected: If, for each specific single configuration of production parameters, only one sound was recorded, then this sound was selected; else the sound with the highest recognition rate, the longest duration and the smallest difference of f_0 intended and f_0 calculated was selected (according to this order). For non-style productions and each vocal effort separately, the sound selection was further limited to f_0 levels for which all vowels investigated were represented. For style productions, the sound selection was generally limited to corresponding style-specific f_0 ranges as practiced by the artist in question.

3. Results

Content: The main body of the database published comprises c. 33800 recordings of sounds of all long Standard German vowels, read texts and songs or arias produced by 16 non-professionals and 24 professionals of straight theatre, contemporary singing styles and European classical singing style, with extensive variation of basic production parameters, including the variation of non-style and style mode for the professionals in terms of separating utterances favouring the intelligibility of vowel quality over sound timbre from utterances focusing on sound aesthetics and standards of a particular speaking or singing style. – The side body of the database presents 830 recordings of sounds of all long Standard German vowels (V context) and of read texts produced with medium vocal effort by 30 native German non-professional reference speakers (Northern part of Switzerland). – The entire corpus thus encompasses c. 34 600 recordings, with sound- and speaker-related information and results of the acoustic analysis.

Presentation, accessibility, terms of use: Sound and speaker information and graphic and numerical display of the acoustic analysis is presented online endued with a graphical user interface and search functionalities [29]. This online website also features an extended description of methodological details, speaker group and single speaker specific details, and the formal conditions for database use. – Database and recordings can be downloaded from the website. However, restrictions apply since the use of the database is limited to scientific purposes only.

Maintenance, future versions: The database will be maintained and corrections will be commented on. Minor changes that do not affect the system of this first version will be assigned with extensions "1.(i)". However, a backup of each existing version will remain accessible in its original form. Future substantial extensions of the database will be labelled accordingly with numbers succeeding "1".

4. Discussion

To the best of our knowledge, this is the first published sound corpus that allows for a direct comparison of acoustic characteristics of vowel sounds for intra- and inter-speaker variation of basic production parameters. The corpus aims at contributing to a phenomenology of the acoustics of the vowel, that is, building up large-scale, language-specific sound descriptions, addressing all variations of production parameters and their possible extension relevant for vowel quality recognition and voice quality classification.

This corpus allows for revisiting basic issues of the acoustics of vowel sounds in terms of a re-evaluation of the acoustic sound characteristics with regard to (1) the performance of acoustic analysis methods, (2) the general variation degree of the vowel quality-related characteristics, (3) phonation type-related differences and similarities (note that whispered and creaky sounds can be compared with voiced sounds, the latter produced on different f_0 levels and with different vocal efforts), (4) vocal effort-related differences and similarities, (5) age- and gender-related differences and similarities (note that speakers of different speaker groups produced sounds at different and similar f_0 levels), (6) f_0-related differences and similarities, (7) speaker-related characteristics, (8) style-related characteristics.

This corpus also contributes to the creation of references needed for future research as empirical basis to test new approaches and models addressing acoustics of vowel and voice quality, including automatic classification and recognition procedures. In this context, future extensions of the database are aimed at in order to increase the number of speakers and to extend the production parameter variation, including additional artistic production styles.

5. Acknowledgements

This work was supported by the Swiss National Science Foundation SNSF Grants No. 100016_143943 and 100016_159350.

6. References

[1] G. G. E. Peterson and H. L. H. Barney, "Control methods used in a study of the vowels," *J. Acoust. Soc. Am.*, vol. 24, no. 2, pp. 175–184, 1952.

[2] J. Hillenbrand and L. Getty, "Acoustic characteristics of American English vowels," *J. Acoust. Soc. Am.*, vol. 97, no. 5, pp. 3099–3111, 1995.

[3] J. Hillenbrand and R. Gayvert, "Identification of steady-state vowels synthesized from the Peterson and Barney measurements," *J. Acoust. Soc. Am.*, vol. 94, no. 2, pp. 668–674, 1993.

[4] S. Zahorian and A. Jagharghi, "Spectral-shape features versus formants as acoustic correlates for vowels," *J. Acoust. Soc. Am.*, vol. 94, no. 4, pp. 1966–82, 1993.

[5] J. M. Hillenbrand, R. A. Houde, and R. T. Gayvert, "Speech perception based on spectral peaks versus spectral shape," *J. Acoust. Soc. Am.*, vol. 119, no. 6, pp. 4041–4054, 2006.

[6] P. F. Assmann and W. F. Katz, "Synthesis fidelity and time-varying spectral change in vowels," *J. Acoust. Soc. Am.*, vol. 117, no. 2, pp. 886–895, 2005.

[7] J. Hillenbrand, "Static and dynamic approaches to vowel perception," in *Vowel Inherent Spectral Change. Modern Acoustics and Signal Processing*, G. Morrison and P. Assman, Eds. Heidelberg: Springer, 2013, pp. 9–30.

[8] G. S. Morrison, "Theories of Vowel Inherent Spectral Change," in *Vowel Inherent Spectral Change. Modern Acoustics and Signal Processing*, G. Morrison and P. Assman, Eds. Heidelberg: Springer, 2013, pp. 31–48.

[9] K. Johnson, "Speaker Normalization in Speech Perception," in *The Handbook of Speech Perception*, D. B. Pisoni and R. E. Remez, Eds. Malden MA: Wiley-Blackwell, 2005, pp. 363–389.

[10] J. M. Hillenbrand and M. J. Clark, "The role of f0 and formant frequencies in distinguishing the voices of men and women," *Atten. Percept. Psychophys.*, vol. 71, no. 5, pp. 1150–1166, 2009.

[11] P. Adank, R. van Hout, R., and R. Smits, ''An acoustic description of the vowels of Northern and Southern Standard Dutch,'' *J. Acoust. Soc. Am.*, vol. 116, no. 3, pp. 1729–1738, 2004.

[12] P. Adank, R. van Hout, R., and R. Smits, ''An acoustic description of the vowels of Northern and Southern Standard Dutch II: regional varieties,'' *J. Acoust. Soc. Am.*, vol. 121, no. 2, pp. 1130–1141, 2007.

[13] C. G. Clopper and D. B. Pisoni, "The Nationwide Speech Project: A new corpus of American English dialects," *Speech Commun.*, vol. 48, no. 6, pp. 633–644, 2006.

[14] K. J. Kallail and F. W. Emanuel, "Formant-frequency differences between isolated whispered and phonated vowel samples produced by adult female subjects," *J. Speech Hear. Res.*, vol. 27, no. 2, pp. 245–251, 1984.

[15] K. J. Kallail and F. W. Emanuel, "An acoustic comparison of isolated whispered and phonated vowel productions," *J. Phon.*, vol. 12, pp. 175–186, 1984.

[16] I. Eklund and H. Traunmüller, "A comparative study of the male and female whispered and phonated versions of the long vowels of Swedish," *TMH-QPSR*, vol. 37, no. 2, pp. 131–134, 1996.

[17] J.-S. Liénard and M.-G. Di Benedetto, "Effect of vocal effort on spectral properties of vowels," *J. Acoust. Soc. Am.*, vol. 106, no. 1, pp. 411–422, Jul. 1999.

[18] H. Traunmüller and A. Eriksson, ''Acoustic effects of variation in vocal effort by men, women and children'' *J. Acoust. Soc. Am.*, vol. 107, no. 6, pp. 3438–3451, 2000.

[19] J. M. Hillenbrand, M. J. Clark, and R. A. Houde, "Some effects of duration on vowel recognition," *J. Acoust. Soc. Am.*, vol. 108, no. 6, pp. 3013–3022, 2000.

[20] J. M. Hillenbrand, M. J. Clark, and T. M. Nearey, "Effects of consonant environment on vowel formant patterns," *J. Acoust. Soc. Am.*, vol. 109, no. 2, pp. 748–763, 2001.

[21] M. Iseli, Y.-L. Shue, and A. Alwan, "Age, sex, and vowel dependencies of acoustic measures related to the voice source," *J. Acoust. Soc. Am.*, vol. 121, no. 4, pp. 2283–2295, 2007.

[22] D. G. Childers and K. Wu, "Gender recognition from speech. Part II: Fine analysis," *J. Acoust. Soc. Am.*, vol. 90, no. 4 Pt 1, pp. 1841–1856, 1991.

[23] J. M. Hillenbrand and R. A. Houde, "A narrow band pattern-matching model of vowel perception," *J. Acoust. Soc. Am.*, vol. 113, no. 2, pp. 1044–1055, 2003.

[24] A. J. S. Ferreira, "Static features in real-time recognition of isolated vowels at high pitch," *J. Acoust. Soc. Am.*, vol. 122, no. 4, pp. 2389–2404, 2007.

[25] P. Forczmański, "Evaluation of Singer's Voice Quality by Means of Visual Pattern Recognition," *J. Voice*, vol. 30, no. 1, p. 127.e21–30, 2016.

[26] J. Sundberg, "Perception of Singing," in *The Psychology of Music*, 3rd ed., D. Deutsch, Ed. San Diego CA: Elsevier, 2013, pp. 69–105.

[27] G. Bloothooft and R. Plomp, "Spectral analysis of sung vowels. I. Variation due to differences between vowels, singers, and modes of singing," *Acoust. Soc. Am.*, vol. 75, no. 4, pp. 1259–1264, 1984.

[28] D. Maurer, P. Mok, D. Friedrichs, and V. Dellwo, "Intelligibility of high-pitched vowel sounds in the singing and speaking of a female Cantonese Opera singer," in *INTERSPEECH 2014 – 15th Annual Conference of the International Speech Communication Association*, September 14–18, Singapore, Proceedings, 2014, pp. 2132–2133.

[29] D. Maurer, Ch. d'Heureuse, H. Suter, V. Dellwo, D. Friedrichs, and T. Kathiresan: "The Zurich Corpus of Vowel and Voice Quality," Version 1.0. Retrieved: http://www.zhcorpus.org [June 14, 2018].

[30] M. Bunch and J. Chapman, "Taxonomy of singers used as subjects in scientific research," *J. Voice*, vol. 14, no. 3, pp. 363–369, 2000.

[31] P. Boersma and D. Weenink: "Praat: doing phonetics by computer," Computer program, Version 6.0.37. Retrieved: http://www.praat.org [March 4, 2018].