Zürcher Hochschule der Künste
Zurich University of the Arts

Universität Zürich[UZH]

Vowel sounds produced with varying production parameters: Conceptualisation and realisation of a database

IAFPA 2016
Annual Conference

Dieter Maurer[1], Thayabaran Kathiresan[2], Volker Dellwo[2]
[1] Institute for the Performing Arts and Film, Zurich University of the Arts, Switzerland
[2] Phonetics Laboratory, Department of Comparative Linguistics, University, University of Zurich, Switzerland
dieter.maurer@zhdk.ch, thayabaran.kathiresan@uzh.ch, volker.dellwo@uzh.ch

Background

With regard to vowel sounds, there is no extensive, empirical database that documents systematic variation of basic production parameters. However, we consider such a corpus to be a prerequisite for a deeper understanding of both phoneme- and speaker-dependent acoustic characteristics.

Concerning phoneme-related aspects, the relation between perceptual vowel recognition, phoneme-specific acoustic characteristics, and variation of basic production parameters such as fundamental frequency (f_o), vocal effort and phonation mode is still a matter of debate [1]. On the one hand, methods to estimate phoneme-specific acoustic characteristics cannot account for all recognisable vowel sounds, above all if f_o exceeds c. 350 Hz [2], or if the phonation is not quasi-periodic (e.g., whisper phonation). On the other hand, recognisable vowel sounds at f_o above their first formant frequency as given in formant statistics for citation-form words contrast with the assumption of a direct relationship between perceived vowel quality and quality-specific formant patterns *F1-F2* or *F1-F2-F3*. This does not only concern the comparison of different speakers but it also holds true for the investigation of sounds of a single speaker [3–6].

Concerning speaker-related aspects, we do not yet know whether existing methods of acoustic analysis are capable of determining speaker-specific acoustic characteristics for voiced sounds produced in isolation and under extensive variation of basic production parameters.

[1] Maurer, D. (2016): Acoustics of the Vowel – Preliminaries. Bern/Frankfurt a.M., Peter Lang.
[2] Traunmüller, H., & Eriksson, A. (1997). A method of measuring formant frequencies at high fundamental frequencies. In Proceedings of Eurospeech (Vol. 97, No. 1, pp. 477-480).
[3] Friedrichs, D., Maurer, D., & Dellwo, V. (2015). The phonological function of vowels is maintained at fundamental frequencies up to 880Hz. The Journal of the Acoustical Society of America, 138 (1), EL36–EL42.
[4] Friedrichs, D., Maurer, D., Suter, H., & Dellwo, V. (2015). Vowel identification at high fundamental frequencies in minimal pairs. In Proceedings of the 18th International Congress of Phonetic Sciences (no. 0434, pp. 1–4).
[5] Maurer, D., & Landis, T. (1996). Intelligibility and spectral differences in high-pitched vowels. Folia Phoniatrica et Logopaedica, 48 (1), 1–10.
[6] Maurer, D., & Landis, T. (2000). Formant pattern ambiguity of vowel sounds. International Journal of Neuroscience, 100 (1–4), 39–76.

Approach

Currently, we are working on a corresponding database for Standard German vowels. In the following, the concept and status of creation of this database, limited to the investigation of untrained speakers, is presented.

Concept

Speaker group A – selection of speakers for the investigation of vowel sounds under extensive variation of basic production parameters: Untrained speakers (children, women, men, gender balanced) are selected according to three qualitative criteria:

- A large vocal range (two octaves for adults and 1.5 octave for children at minimum), in order to allow for the investigation of extensive f_o variation in vowel sound production and perception
- The ability to reproduce sounds on a given f_o (presented as a piano sound) in order to allow for a systematic comparison of all sounds investigated
- Clear vowel articulation for a range of f_o of one octave at minimum, in order to allow for an investigation of vowel recognition related to very different f_o levels

Sound production: Each speaker produces the sounds of the long Standard German vowels /i-y-e-ø-ɛ-a-o-u/ and varies basic production parameters such as

- f_o (C-major scale up and down the entire vocal range; pitches presented to the speaker as electronic piano sounds)
- vocal effort (medium, low, high, shouted)
- phoneme context (V for all f_o and all vocal efforts, sVsV for middle and high f_o and medium vocal effort only)
- phonation mode (breathy in V context, whisper in V and sVsV context)

Each speaker also reads a phonetically balanced text.

If, during a recording session, the speaker or examiner believes that a sound could possibly be improved concerning articulation and vowel recognition, additional productions are recorded for the corresponding production condition.

Speaker group B – selection of speakers for further references of vowel sounds: Untrained adults (gender balanced) are selected who are able to clearly articulate vowel sounds for a range of f_o of one octave.

Sound production: Each speaker produces voiced sounds of the above German vowels at three levels of f_o, i.e. 220–262–440Hz for women, and 131–220–262Hz for men, with medium vocal effort. The lower level of f_o corresponds to average f_o as given in formant statistics for vowel sounds in citation-form words, the other levels of f_o allow for comparisons of sounds produced by speakers different in age and/or gender.

Each speaker also reads a phonetically balanced text.

These sounds serve as a reference for the investigation of speakers of group A producing sounds under extensive variation of production parameters: Given the f_o levels of the reference sounds and medium vocal effort in sound production, it can be tested whether the spectral characteristics of the sounds of the speakers of both speaker groups A and B correspond, or whether they differ.

Concept (continued)

Recording: All sounds are digitally recorded in a quiet room and with a constant speaker-microphone distance of 30 cm. The microphone input gain is adjusted according to the vocal effort investigated. In order to subsequently determine the actual sound pressure level, for each recording session, a 1 kHz sinus wave is recorded with a -20 dB gain using a specific calibration tool.

Editing: Each sound is annotated with standard information on speaker, production condition and recording parameters.

Acoustic analysis: In order to provide additional indicative information, the database also features the results of acoustic analysis for the vowel nuclei (average f_o and f_o contour, average formant frequencies, formant tracks and spectrogram, average spectrum, LPC filter curve for the middle window of analysis).

Vowel recognition – listening test: For further indicative information, the results of a listening test (assignment of the perceived vowel quality) involving five professionally trained singers and speakers are given.

If more than one sound is recorded for a given configuration of production parameters, the sound with the highest identification score is selected for the database.

Actual Status

Speaker group A: Up to now, we have recorded eight adults and eight children (gender balanced). With the exception of one child, all speakers produced the sounds over a vocal range of two octaves or more. Considering that there is a vowel sound for each production condition (sound selection according to the highest identification score of the listening test), depending on individual vocal ranges, 450–550 systematic recordings are available for each single speaker, or c. 8000 recordings for all 16 speakers.

Speaker group B: In parallel, we have recorded 20 adults (gender balanced), each producing 24 vowel sounds and reading a text, or 500 recordings for all 20 speakers.

Additions

We are also recording the sounds of professional singers and actresses and actors, including additional variations of production parameters, such as production style, phoneme context in minimal pairs, and creak phonation. Above all, this investigation allows for a comparison of trained and untrained speakers, different artistic speaking/singing styles, and vowel acoustics/recognition on very high pitches.

Relevance to forensic phonetics

Although this database was collected under very controlled laboratory settings, it should be helpful to understand phoneme- and speaker-specific characteristics under strong variability in vocalic utterances, e.g., concerning

- Speech expressing strong emotions
- Speech habits including large variations of f_o often combined with register changes
- Vocalic shouts

In future work, we are planning to test automatic and human recognition of speakers based on vowel sounds produced over an extensive range of f_o (up to c. 880 Hz) after initial training at fundamental frequencies typical in citation form (c. 220Hz in females, c. 131Hz in males). Since the vocal tract is increasingly undersampled with higher f_o it is unclear to what degree speaker specific characteristics of the vocal tract are still present in these signals. Results may have strong implications on forensic settings where acoustic trace and comparison material differs substantially in fundamental frequency.

Examples (provided online)

The following illustrations are given online:

Figure 1: Sound sample of a single female speaker in the database; matrix according to the production parameters investigated.

Figure 2: Illustration of standard information, of acoustics analysis and of the indications on vowel recognition (result of the listening test) given for a single sound.

Figure 3: Illustration of three sound series of the vowels /u, y, i/ produced by a female speaker with medium vocal effort over a range of f_o of 220–880 Hz.

Figure 4: Comparison of sounds of /o/ produced by a child, a woman and a man at comparable levels of f_o; illustration of a corresponding decrease or disappearance of expected age- and gender-related spectral characteristics < 2 kHz.

Figure 5: Comparison of whispered and voiced sounds of /o/ produced by a woman, the voiced sounds including variation of f_o.

Figure 6: Pitch-contours of speech with large variations of f_o; examples to experience in everyday life (additional illustration; not part of the database).

Please refer to http://www.phones-and-phonemes.org/IAFPA2016.