The Zurich Corpus of Vowel and Voice Quality, Beta Version

zhcorpus.org

(Last update 2019-02-10)

Dieter Maurer, Christian d'Heureuse, Heidy Suter, Volker Dellwo, Daniel Friedrichs, Thayabaran Kathiresan

Handbook

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 in straight theatre and singers in the styles of contemporary singing, and European classical singing). The database focuses on the investigation 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 were also elicited. The database is accessible for scientific use and further extensions are in progress.

1 Overview (Interspeech 2018 paper)

D. Maurer, C. d'Heureuse, H. Suter, V. Dellwo, D. Friedrichs, T. Kathiresan, "The Zurich Corpus of Vowel and Voice Quality, Version 1.0," in *INTERSPEECH 2018 – 98th Annual Conference of the International Speech Communication Association, September 2-6, Hyderabad, India, Proceedings*, 2018, 1417–1421.

Paper (PDF)

2 Terms and abbreviations

2.1 Entire listing

Below, all abbreviations used here are listed in alphabetical order:

b	breathy phonation
c	creaky phonation (vocal fry)
CS	contemporary singing style (generic term, including the substyles contemporary musical theatre, pop, and jazz)
D	standard recording, but deficient (details see individual sound comments)
EC	European classical singing style
f_o	fundamental frequency
F_1, F_1, ...	formant frequencies
hgh	high vocal effort
low	low vocal effort
m	male
med	medium vocal effort
ms	musical scale
N	non-style
ns	not specified
rp	reference pitch
S	standard recording
sh	shouted
ST	straight theatre speaking style
sVsV	vowel sound produced in s-Vowel-s-Vowel context
//Text//	text read or text sung (indication for speech identity)
tr	text read
ts	text sung
v	voiced
V	vowel sound produced in isolation
var	varied vocal effort
w	female
w	whispered

2.2 Speakers and styles

The database presented includes recordings of four speaker subgroups: untrained and nonprofessional speakers (N), professional straight theatre actresses and actors (ST), singers of contemporary singing style (CS, substyles include contemporary musical theatre, pop, and jazz), and singers of European classical singing (EC):

N	non-style
ST	straight theatre speaking style
CS	contemporary singing style
EC	European classical singing style

Note: In this handbook, both speakers and singers are referred to as speaker(s) as a generic term.

For details of speakers and utterances, see the Method section.

2.3 Production parameters and sound status

The database presents recordings of utterances produced with varying basic production parameters such as phonation type, vocal effort, fundamental frequency, vowel context and speaking or singing style. The corresponding abbreviations are used for the different parameters:

Abbreviations for phonation types:

v	voiced
b	breathy
c	creaky (vocal fry)
w	whispered

Abbreviations for vocal effort:

med	medium
low	low
hgh	high
var	varying vocal effort

Abbreviations for production modes:

tr	text read
ts	text sung
rp	f_o according to reference pitches given
ms	f_o according to musical scale (C-Major)
sh	shouted
ns	not specified (no specific production mode)

Abbreviations for styles, see Chapter 2.2.

Abbreviations for vowel context:

V	in isolation
sVsV	in s-V-s-V context

Abbreviations for sound status:

S	standard recording
D	standard recording, but deficient (details see individual sound comments)

3 Rationale

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_0 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 nonprofessional speakers and trained and professionally active speakers and singers.

The sound corpus thus allows for a direct comparison of acoustic characteristics of vowel sounds for intra- and inter-speaker variation of basic production parameters (for details, see the Method section.) 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 basic variations of production parameters and their possible relevance for vowel quality recognition and voice quality classification.

The phenomenological approach of a large-scale sound documentation adopted here has some characteristics which differ from many experimental laboratory studies. Above all, as described in details below, it includes varying recording room conditions, substantial variations in sound duration, and a small number of professionally trained speakers or singers as listeners for the vowel recognition tests, including two of the authors as listeners.

4 Details of method

4.1 Speakers and utterances

Speakers of extensive investigation (main body): As shown in Table 1 (see speaker groups A to D), 16 nonprofessionals (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). – Nonprofessionals were selected according to two criteria: a minimal vocal range for vowel production of 2 octaves (24 semitones) for adults and 1.5 octaves (19 semitones) for children, with vowels recognisable over a range of 1 octave in minimum for both adults and children. The children's age was set at 7 to 10 years to control for voice changes that happen in puberty. Children – and boys in particular – below 10 years of age are at a stage well before that of undergoing any voice changes. The minimal age of 7 was set to ensure that the children are mature enough to follow instructions correctly and are equipped with the physical and mental stamina and attention span needed for the recording sessions. 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), who are not native speakers of German but who perform on stage professionally 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**		**Adults**		**Speakers**
	f	**m**	**f**	**m**	**total**
Main body					
A : Nonprofessionals	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 : Nonprofessionals	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_o (monotonous pitch levels according to C-major scale, full f_o range according to assigned register/voice range). All utterances were made by the speakers as nonprofessional (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 voiced sounds for the same set of long Standard German vowels in their own respective artistic style of singing or speaking with the above mentioned production parameter variation for a range of f_o that reflects their artistic style (for details, see below, the Notes of this chapter). 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_o range (≥ 523 Hz for children and women, ≥ 330 Hz for tenors and high male voices, ≥ 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. For two children, additional songs were also recorded.
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_o 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 fo
E				

Reference speakers (side body): Vowel production with very limited f_o variation of 30 native Swiss German nonprofessional speakers (10 children, aged 7–9, and 20 adults, aged 18–52, gender balanced) were also investigated (see Table 1, group E). The speakers were selected according to their native Swiss German dialect in the Northern part of Switzerland, their command of speaking Standard German (primary teaching language in Switzerland's schools) and their ability to produce vowel sounds on a specific pitch over an f_o 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_o 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 pitch contour of real speech prosody and to allow for a comparison of sounds on different and similar f_o for the different age and gender subgroups. In addition, the speakers read the same phonetically balanced text as the nonprofessional and professional speakers . – 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.

4.2 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 high-frequency condenser microphone (Sennheiser MKH 40P 48, cardioid characteristics) 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 sinus 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. – 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 when investigating the vocal range limits. – If either speaker or investigator believed that a repeated recording of a certain utterance would improve sound or vowel quality, the recording was repeated one or several more times.
f_o scale: For each speaker, a comfortable "middle" pitch 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" pitch 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 via digital piano sound or was vocally presented by the investigator.
Corrections: Upon successful recording and editing of the recorded sounds, a listening test (see below) was performed for each speaker. If the listening test did not yield satisfactory results (low recognition rate, short duration), the speaker was asked to come in for another recording session to see whether vowel production could be improved upon. These types of corrections, however, were limited to non-style productions only, and they were not feasible for all speakers due to scheduling problems 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.

4.3 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=25ms, time steps=5ms, pre-emphasis=50Hz) 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 LTAS (0–5.5 and 0–11.1 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, 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 are overlaid onto the spectrum in order to illustrate the correspondence between spectral peaks and calculated formants. For texts and songs/arias, graphic representation included the display of the sound wave, the f_o contour, the spectrogram and the LTAS. – Numerical average values of f_o and formant patterns were added to the sound information.
Crosschecks: All sounds were acoustically crosschecked, and sounds with marked background noise not related to vowel sound production were removed. Graphic representations 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, in most cases relating to C-major scale.

4.4 Listening test

Listeners: Five phonetic expert listeners (professionally trained singers or actors or voice teachers) performed listening tests for all vowel sounds. 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: the listeners were asked to listen to each sound and to assign this sound to what they believe the correspondent vowel quality to be. The answer options to choose from are: (1) a single specific Standard German vowel (/i–y–e–ø–ɛ–a–ɔ–o–u/), (2) a vowel boundary region of two vowels maximum, (3) "no vowel" or (4) a free comment. If a sound was difficult to identify they could listen to it repeatedly by means of a "repeat play" button. – 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. – For the screen used by the listeners when testing the vowel quality, see Appendix A3.

4.5 Sounds recorded and sounds selected for publication

For the production parameters documented in this first version of the corpus, in total, c. 56'600 recordings were made for all 70 speakers. As mentioned, in many cases, two or multiple recordings were made for a specific configuration of production parameters in order to obtain the best possible vowel or sound quality. For the publication of the open-access database, a subset of the recorded sounds was selected: If only one sound was recorded for a specific configuration of production parameters, then this sound was selected; if there are multiple recordings for one specific parameter, the sound with the highest recognition rate, the longest duration and the smallest difference of f_o intended and f_o calculated was selected (according to this order). For non-style productions and each vocal effort separately, the sound selection was further limited an f_o range in which all vowels investigated were represented. For example, if at a very low or high pitch, an [i:] could not be produced, then this pitch was not included into the production matrix, even though all of the other long vowels were produced successfully. For style productions, the sound selection was generally limited to corresponding style-specific f_o ranges as practiced by the artist in question.

Additional information concerning the sound selection:

- As mentioned, sounds were selected according to a ranking.
- Thereby, a ranking level relates always to a given single configuration of production parameters, represented in a single position in the production matrix; the ranking thus concerns sounds of such a single configuration.
- According to the above ranking procedure, the following ranking levels occurred:
    - r0 = sounds not included in the listening test; this ranking concerns all read texts and all songs. In addition, sounds which are under correction are also assigned with r0, until the corrections are made.
    - r1, r2, r3, r4 for sound(s) related to single production parameters including sound(s) with recognised vowel quality:
        - r1 = the sound with highest vowel recognition rate, longest duration and smallest difference of f_o intended and f_o calculated (in this order) compared to the other sounds; note that if only one sound relates to a single production parameter

configuration and the vowel quality is recognised by a majority of the listeners, automatically, r1 level is assigned
        - r2 = sound(s) with the same vowel recognition rate than the sound ranked as r1, but shorter duration and larger difference of f_o intended and f_o calculated (in this order)
        - r3 = sound(s) with a lower vowel recognition rate than the sound ranked as r1
        - r4 = sound(s) with no vowel recognition
    - r5, r6, r7 for sound(s) related to single production parameters with no sound(s) with recognised vowel quality:
        - r5 = a single sound with no vowel recognition related to single production parameter configuration
        - r6 = a selected sound of a sound sample with no vowel recognition related to single production parameter configuration (all sounds related to that configuration ranked as r7, and one sound manually selected for publication and correspondingly assigned as r6)
        - r7 = other sound(s) of a sound sample with no vowel recognition related to single production parameter configuration
- In this first version of the database, sounds with ranking levels r0, r1, r5 and r6 were selected in order to provide a systematic sound configuration. This means that for each speaker, each style and each single production parameter configuration, one sound is presented.

5 Database content I: General characteristics, form of online presentation, and quicklinks

The main body of the database published comprises c. 33'800 recordings of sounds of all long Standard German vowels, read texts and songs or arias produced by 12 nonprofessionals (adults and children, gender balanced) and 32 professionals of straight theatre, contemporary singing styles and European classical singing style (gender balanced), 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 professional speaking or singing style. – The side body of the database presents 830 recordings of sounds of all long Standard German vowels (V context, medium vocal effort) and of read texts produced by 30 native German nonprofessional reference speakers (Northern part of Switzerland) with medium vocal effort. – The entire corpus thus encompasses c. 34'600 recordings, with sound- and speaker-related information and results of the acoustic analysis.

In this online presentation, sound and speaker information and graphic and numerical display of the acoustic analysis is given endued with a graphical user interface and search functionalities (see the corresponding descriptions in the Help menu of the sound archive).

Database and recordings can be downloaded from the homepage of this website. However, restrictions apply since the use of the database is limited to scientific purposes only (for details see Chapter 8).

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".

For an overview over the entire sample of recorded sounds, use the following two links:

Overview of the entire sound sample (production matrix)

Overview of the entire sound sample (list of texts and vowel sounds in V and sVsV context)

In the following chapters, speaker and speaker group specific links are given including details regarding gender, age and f_o ranges.

6 Database content II: Details of the main body

6.1 Nonprofessional speakers

6.1.1 Children

6.1.1.1 Female

Female, aged 8, speaker ID 1009
Vocal range documented = G_3–A_5, 196–880 Hz
Overview of sound sample (539 sounds)

Female, aged 10, speaker ID 1034
Vocal range documented = F_3–A_5, 175–880 Hz
Overview of sound sample (498 sounds)

Female, aged 8, speaker ID 1037
Vocal range documented = G_3–A_5, 196–880 Hz
Overview of sound sample (514 sounds)

Female, aged 10, speaker ID 1038
Vocal range documented = G_3–G_5, 196–784 Hz
Overview of sound sample (420 sounds)

6.1.1.2 Male

Male, aged 8, speaker ID 1054
Vocal range documented = G_3–F_5, 196–698 Hz
Overview of sound sample (434 sounds)

Male, aged 7, speaker ID 1056
Vocal range documented = F_3–A_5, 175–880 Hz
Overview of sound sample (474 sounds)

Male, aged 8, speaker ID 1057
Vocal range documented = H_3–G_5, 247–784 Hz
Overview of sound sample (394 sounds)

Male, aged 8, speaker ID 1058
Vocal range documented = A_3–G_5, 220–784 Hz
Overview of sound sample (434 sounds)

6.1.2 Adult speakers

6.1.2.1 Women

Female, aged 35, speaker ID 1027
Vocal range documented = C_3–A_5, 131–880 Hz
Overview of sound sample (530 sounds)

Female, aged 34, speaker ID 1036
Vocal range documented = C_3–A_5, 131–880 Hz
Overview of sound sample (578 sounds)

Female, aged 25, speaker ID 1039
Vocal range documented = G_3–C_6, 196–1047 Hz
Overview of sound sample (530 sounds)

Female, aged 28, speaker ID 1088
Vocal range documented = G_3–A_5, 196–880 Hz
Overview of sound sample (482 sounds)

6.1.2.1 Men

Male, aged 23, speaker ID 1044
Vocal range documented = F_2–G_5, 87–784 Hz
Overview of sound sample (690 sounds)

Male, aged 31, speaker ID 1045
Vocal range documented = A_2–E_5, 110–659 Hz
Overview of sound sample (594 sounds)

Male, aged 40, speaker ID 1051
Vocal range documented = G_2–E_5, 98–659 Hz
Overview of sound sample (594 sounds)

Male, aged 24, speaker ID 1063
Vocal range documented = G_2–G_5, 98–784 Hz
Overview of sound sample (682 sounds)

6.1.3 All nonprofessional speakers

Children, aged 7 to 10, speaker ID's = 1009, 1034, 1037, 1038, 1054, 1056, 1057, 1058
Vocal range documented = F_3–A_5, 175–880 Hz
Overview of sound sample (3707 sounds)

Women, aged 25 to 35, speaker ID's = 1027, 1036, 1039, 1088
Vocal range documented = C_3–C_6, 131–1047 Hz
Overview of sound sample (2120 sounds)

Men, aged 23 to 40, speaker ID's = 1044, 1045, 1051, 1063
Vocal range documented = F_2–G_5, 87–784 Hz
Overview of sound sample (2560 sounds)

All nonprofessional speakers, aged 7 to 40, speaker ID's = 1009, 1027, 1034, 1036, 1037, 1038, 1039, 1044, 1045, 1051, 1054, 1056, 1057, 1058, 1063, 1088
Vocal range documented = F_2–C_6, 87–1047 Hz
Overview of sound sample (8395 sounds)

6.2 ST speakers

6.2.1 Women

Female, aged 44, speaker ID = 1046
Professional ranking = 2 (2-BAA)
Total number of sounds = 1059
Vocal range documented for non-style productions = E_3–C_6, 165–1047 Hz
Overview of sound sample for non-style productions (561 sounds)
Vocal range documented for style productions = E_3–C_6, 165–1047 Hz
Overview of sound sample for style productions (498 sounds)
Vocal range documented for all productions = E_3–C_6, 165–1047 Hz
Overview of entire sound sample (1059 sounds)

Female, aged 32, speaker ID = 1048
Professional ranking = 2 (2-BAA)
Total number of sounds = 923
Vocal range documented for non-style productions = C_3–G_5, 131–784 Hz
Overview of sound sample for non-style productions (521 sounds)
Vocal range documented for style productions = D_3–G_5, 147–784 Hz
Overview of sound sample for style productions (402 sounds)
Vocal range documented for all productions = C_3–G_5, 131–784 Hz
Overview of entire sound sample (923 sounds)

Female, aged 51, speaker ID = 1052
Professional ranking = 2 (2-BAA)
Total number of sounds = 1163
Vocal range documented for non-style productions = H_2–C_6, 123–1047 Hz
Overview of sound sample for non-style productions (665 sounds)
Vocal range documented for style productions = C_3–A_5, 131–880 Hz
Overview of sound sample for style productions (498 sounds)
Vocal range documented for all productions = H_2–C_6, 123–1047 Hz
Overview of entire sound sample (1163 sounds)

Female, aged 34, speaker ID = 1053
Professional ranking = 2 (2-BAA)
Total number of sounds = 899
Vocal range documented for non-style productions = E_3–G_5, 165–784 Hz
Overview of sound sample for non-style productions (481 sounds)
Vocal range documented for style productions = E_3–G_5, 165–784 Hz
Overview of sound sample for style productions (418 sounds)
Vocal range documented for all productions = E_3–G_5, 165–784 Hz
Overview of entire sound sample (899 sounds)

6.2.2 Men

Male, aged 39, speaker ID = 1003
Professional ranking = 2 (2-BAB) and T-AA
Total number of sounds = 1083
Vocal range documented for non-style productions = E_2–D_5, 82–587 Hz
Overview of sound sample for non-style productions (585 sounds)
Vocal range documented for style productions = G_2–C_5, 98–523 Hz
Overview of sound sample for style productions (498 sounds)
Vocal range documented for all productions = E_2–D_5, 82–587 Hz
Overview of entire sound sample (1083 sounds)

Male, aged 43, speaker ID = 1047
Professional ranking = 2 (2-BAA)
Total number of sounds = 1171
Vocal range documented for non-style productions = E_2–E_5, 82–659 Hz
Overview of sound sample for non-style productions (601 sounds)
Vocal range documented for style productions = E_2–F_5, 82–698 Hz
Overview of sound sample for style productions (570 sounds)
Vocal range documented for all productions = E_2–F_5, 82–698 Hz
Overview of entire sound sample (1171 sounds)

Male, aged 26, speaker ID = 1049
Professional ranking = 2 (2-BAA)
Total number of sounds = 1227
Vocal range documented for non-style productions = D_2–E_5, 73–659 Hz
Overview of sound sample for non-style productions (681 sounds)
Vocal range documented for style productions = F_2–E_5, 87–659 Hz
Overview of sound sample for style productions (546 sounds)
Vocal range documented for all productions = D_2–E_5, 73–659 Hz
Overview of entire sound sample (1227 sounds)

Male, aged 32, speaker ID = 1050
Professional ranking = 2 (2-BAA)
Total number of sounds = 1307
Vocal range documented for non-style productions = C_2–E_5, 65–659 Hz
Overview of sound sample for non-style productions (713 sounds)
Vocal range documented for style productions = E_2–E_5, 82–659 Hz
Overview of sound sample for style productions (594 sounds)
Vocal range documented for all productions = C_2–E_5, 65–659 Hz
Overview of entire sound sample (1307 sounds)

6.2.3 All speakers.

Women, aged 32 to 51, speaker ID's = 1046, 1048 , 1052, 1053
Vocal range documented for all productions = H_2–C_6, 123–1047 Hz
Overview of entire sound sample (4044 sounds)

Men, aged 26 to 43, speaker ID's = 1003, 1047, 1049, 1050
Vocal range documented = C_2–F_5, 65–698 Hz
Overview of entire sound sample (4788 sounds)

All ST speakers, aged 26 to 51, speaker ID's = 1003, 1046, 1047, 1048, 1049, 1050, 1052, 1053
Vocal range documented = C_2– C_6, 65–1047 Hz
Overview of entire sound sample (8832 sounds)

6.3 CS speakers

6.3.1 Women

Female, aged 26, speaker ID = 1001
Professional ranking = 2 (2-BBAA)
Total number of sounds = 1003
Vocal range documented for non-style productions = D_3–A_5, 147–880 Hz
Overview of sound sample for non-style productions (553 sounds)
Vocal range documented for style productions = G_3–C_6, 196–1047 Hz
Overview of sound sample for style productions (450 sounds)
Vocal range documented for all productions = D_3–C_6, 147–1047 Hz
Overview of entire sound sample (1003 sounds)

Female, aged 46, speaker ID = 1006
Professional ranking = 2 (2-BC) and T-AA
Total number of sounds = 915
Vocal range documented for nonstyle productions = A_2–A_5, 110–880 Hz
Overview of sound sample for non-style productions (553 sounds)
Vocal range documented for style productions = F_3–F_5, 175–698 Hz
Overview of sound sample for style productions (362 sounds)
Vocal range documented for all productions = A_2–A_5, 110–880 Hz
Overview of entire sound sample (915 sounds)

Female, aged 34, speaker ID = 1023
Professional ranking = 2 (2-BBAA)
Total number of sounds = 1109
Vocal range documented for non-style productions = D_3–C_6, 147–1047 Hz
Overview of sound sample for non-style productions (649 sounds)
Vocal range documented for style productions = F_3–A_5, 175–880 Hz
Overview of sound sample for style productions (460 sounds)
Vocal range documented for all productions = D_3–C_6, 147–1047 Hz
Overview of entire sound sample (1109 sounds)

Female, aged 50, speaker ID = 1031
Professional ranking = 2 (2-BBAA)
Total number of sounds = 1035
Vocal range documented for non-style productions = E_3–C_6, 165–1047 Hz
Overview of sound sample for non-style productions (545 sounds)
Vocal range documented for style productions = G_3–C_6, 196–1047 Hz
Overview of sound sample for style productions (490 sounds)
Vocal range documented for all productions = E_3–C_6, 165–1047 Hz
Overview of entire sound sample (1035 sounds)

6.3.2 Men

Male, aged 29, speaker ID = 1002
Professional ranking = 2 (2-BBAA/2-BBB)
Total number of sounds = 1150
Vocal range documented for nonstyle productions = D_2–F_5, 73–698 Hz
Overview of sound sample for non-style productions (649 sounds)
Vocal range documented for style productions = G_2–C_5, 98–523 Hz
Overview of sound sample for style productions (501 sounds)
Vocal range documented for all productions = D_2–F_5, 73–698 Hz
Overview of entire sound sample (1150 sounds)

Male, aged 27, speaker ID = 1030
Professional ranking = 2 (2-BBAA)
Total number of sounds = 1053
Vocal range documented for non-style productions = G_2–D_5, 98–587 Hz
Overview of sound sample for non-style productions (593 sounds)
Vocal range documented for style productions = A_2–D_5, 110–587 Hz
Overview of sound sample for style productions (460 sounds)
Vocal range documented for all productions = G_2–D_5, 98–587 Hz
Overview of entire sound sample (1053 sounds)

Male, aged 28, speaker ID = 1033
Professional ranking = 2 (2-BBAA)
Total number of sounds = 1140
Vocal range documented for non-style productions = A_2–G_5, 110–784 Hz
Overview of sound sample for non-style productions (633 sounds)
Vocal range documented for style productions = A_2–E_5, 110–659 Hz
Overview of sound sample for style productions (507 sounds)
Vocal range documented for all productions = A_2–G_5, 110–784 Hz
Overview of entire sound sample (1140 sounds)

Male, aged 32, speaker ID = 1064
Professional ranking = 2 (2-BBAA)
Total number of sounds = 1259
Vocal range documented for non-style productions = D_2–F_5, 73–698 Hz
Overview of sound sample for non-style productions (713 sounds)
Vocal range documented for style productions = G_2–E_5, 98–659 Hz
Overview of sound sample for style productions (546 sounds)
Vocal range documented for all productions = D_2–F_5, 73–698 Hz
Overview of entire sound sample (1259 sounds)

6.3.3 All speakers

Women, aged 26 to 50, speaker ID's = 1001, 1006, 1023, 1031
Vocal range documented = A_2–C_6, 110–1047 Hz
Overview of entire sound sample (4062 sounds)

Men, aged 27 to 32, speaker ID's = 1002, 1030, 1033, 1064
Vocal range documented = D_2–G_5, 73–784 Hz
Overview of entire sound sample (4602 sounds)

All CS speakers, aged 26 to 51, speaker ID's = 1001, 1002, 1006, 1023, 1030, 1031, 1033, 1064
Vocal range documented = D_2–C_6, 73–1047 Hz
Overview of entire sound sample (8664 sounds)

6.4 EC speakers

6.4.1 Women

Female, aged 54, speaker ID = 1004
Professional ranking = 2 (2-BCB) and T-AA
Total number of sounds = 1028
Vocal range documented for nonstyle productions = D_3–C_6, 147–1047 Hz
Overview of sound sample for non-style productions (625 sounds)
Vocal range documented for style productions = A_3–A_5, 220–880 Hz
Overview of sound sample for style productions (403 sounds)
Vocal range documented for all productions = D_3–C_6, 147–1047 Hz
Overview of entire sound sample (1028 sounds)

Female, aged 56, speaker ID = 1005
Professional ranking = 2 (2-BCB) and T-AA
Total number of sounds = 958
Vocal range documented for nonstyle productions = H_2–A_5, 123–880 Hz
Overview of sound sample for non-style productions (545 sounds)
Vocal range documented for style productions = A_3–A_5, 220–880 Hz
Overview of sound sample for style productions (413 sounds)
Vocal range documented for all productions = 123–880 Hz
Overview of entire sound sample (958 sounds)

Female, aged 41, speaker ID = 1032
Professional ranking = 2 (2-BCAA)
Total number of sounds = 1061
Vocal range documented for nonstyle productions = C_3–C_6, 131–1047 Hz
Overview of sound sample for non-style productions (633 sounds)
Vocal range documented for style productions = C_4–C_6, 262–1047 Hz
Overview of sound sample for style productions (428 sounds)
Vocal range documented for all productions = C_3–C_6, 131–1047 Hz
Overview of entire sound sample (1061 sounds)

Female, aged 30, speaker ID = 1102
Professional ranking = 2 (2-BCAA)
Total number of sounds = 1069
Vocal range documented for nonstyle productions = C_3–C_6, 131–1047 Hz
Overview of sound sample for non-style productions (641 sounds)
Vocal range documented for style productions = C_4–C_6, 262–1047 Hz
Overview of sound sample for style productions (428 sounds)
Vocal range documented for all productions = C_3–C_6, 131–1047 Hz
Overview of entire sound sample (1069 sounds)

6.4.2 Men

Male, aged 44, speaker ID = 1007
Professional ranking = 2 (2-ACAA) and T-AA
Total number of sounds = 972
Vocal range documented for nonstyle productions = D_2–H_4, 73–494 Hz
Overview of sound sample for non-style productions (577 sounds)
Vocal range documented for style productions = G_2–G_4, 98–392 Hz
Overview of sound sample for style productions (395 sounds)
Vocal range documented for all productions = D_2–H_4, 73–494 Hz
Overview of entire sound sample (972 sounds)

Male, aged 29, speaker ID = 1042
Professional ranking = 2 (2-ACAA)
Total number of sounds = 876
Vocal range documented for nonstyle productions = H_2–C_5, 123–523 Hz
Overview of sound sample for non-style productions (465 sounds)
Vocal range documented for style productions = C_3–C_5, 131–523 Hz
Overview of sound sample for style productions (411 sounds)
Vocal range documented for all productions = H_2–C_5, 123–523 Hz
Overview of entire sound sample (876 sounds)

Male, aged 25, speaker ID = 1060
Professional ranking = 2 (2-BCAA)
Total number of sounds = 948
Vocal range documented for nonstyle productions = F_2–C_5, 87–523 Hz
Overview of sound sample for non-style productions (537 sounds)
Vocal range documented for style productions = G_2–A_4, 98–440 Hz
Overview of sound sample for style productions (411 sounds)
Vocal range documented for all productions = F_2–C_5, 87–523 Hz
Overview of entire sound sample (948 sounds)

Male, aged 31, speaker ID = 1103
Professional ranking = 2 (2-ACAA) and T-AA
Total number of sounds = 916
Vocal range documented for nonstyle productions = G_2–C_5, 98–523 Hz
Overview of sound sample for non-style productions (505 sounds)
Vocal range documented for style productions = C_3–C_5, 131–523 Hz
Overview of sound sample for style productions (411 sounds)
Vocal range documented for all productions = A_2–C_5, 110–523 Hz
Overview of entire sound sample (916 sounds)

6.4.3 All speakers

Women, aged 30 to 56, speaker ID's = 1004, 1005, 1032, 1102
Vocal range documented = H_2–C_6, 123–1047 Hz
Overview of entire sound sample (4116 sounds)

Men, aged 25 to 44, speaker ID's = 1007, 1042, 1060, 1103
Vocal range documented = D_2–C_5, 73–523 Hz
Overview of entire sound sample (3712 sounds)

All EC speakers, aged 25 to 56, speaker ID's = 1004, 1005, 1007, 1032, 1042, 1060, 1102, 1103
Vocal range documented =D_2–C_6, 73–1047 Hz
Overview of entire sound sample (7828 sounds)

7 Database content III: Details of the side body

7.1 Reference group

7.1.1 Children

7.1.1.1 Female

Female, aged 8, speaker ID 1095
Vocal range documented = 15 semitones, A_3–C_5, 220–523 Hz
Overview of sound sample (33 sounds)

Female, aged 9, speaker ID 1096
Vocal range documented = 15 semitones, A_3–C_5, 220–523 Hz
Overview of sound sample (33 sounds)

Female, aged 8, speaker ID 1097
Vocal range documented = 15 semitones, A_3–C_5, 220–523 Hz
Overview of sound sample (33 sounds)

Female, aged 8, speaker ID 1098
Vocal range documented = 15 semitones, A_3–C_5, 220–523 Hz
Overview of sound sample (33 sounds)

Female, aged 7, speaker ID 1101
Vocal range documented = 15 semitones, A_3–C_5, 220–523 Hz
Overview of sound sample (33 sounds)

7.1.1.2 Male

Male, aged 9, speaker ID 1089
Vocal range documented = 15 semitones, A_3–C_5, 220–523 Hz
Overview of sound sample (33 sounds)

Male, aged 8, speaker ID 1090
Vocal range documented = 15 semitones, A_3–C_5, 220–523 Hz
Overview of sound sample (33 sounds)

Male, aged 8, speaker ID 1091
Vocal range documented = 15 semitones, A_3–C_5, 220–523 Hz
Overview of sound sample (33 sounds)

Male, aged 8, speaker ID 1092
Vocal range documented = 15 semitones, A_3–C_5, 220–523 Hz
Overview of sound sample (33 sounds)

Male, aged 7, speaker ID 1093
Vocal range documented = 15 semitones, A_3–C_5, 220–523 Hz
Overview of sound sample (33 sounds)

7.1.1.3 All speakers

Female children, aged 7 to 9, speaker ID's = 1095, 1096, 1097, 1098, 1101
Vocal range documented = 15 semitones, A_3–C_5, 220–523 Hz
Overview of sound sample (165 sounds)

Male children, aged 8 to 9, speaker ID's =1089, 1090, 1091, 1092, 1093
Vocal range documented = 15 semitones, A_3–C_5, 220–523 Hz
Overview of sound sample (165 sounds)

All children, aged 7 to 9, speaker ID's = 1095, 1096, 1097, 1098, 1101, 1089, 1090, 1091, 1092, 1093
Vocal range documented = 15 semitones, A_3–C_5, 220–523 Hz
Overview of sound sample (330 sounds)

7.1.2 Adults

7.1.2.1 Women

Female, aged 26, speaker ID 1012
Vocal range documented = 12 semitones, A_3–A_4, 220–440 Hz
Overview of sound sample (25 sounds)

Female, aged 24, speaker ID 1016
Vocal range documented = 12 semitones, A_3–A_4, 220–440 Hz
Overview of sound sample (25 sounds)

Female, aged 20, speaker ID 1020
Vocal range documented = 12 semitones, A_3–A_4, 220–440 Hz
Overview of sound sample (25 sounds)

Female, aged 26, speaker ID 1021
Vocal range documented = 12 semitones, A_3–A_4, 220–440 Hz
Overview of sound sample (25 sounds)

Female, aged 28, speaker ID 1040
Vocal range documented = 12 semitones, A_3–A_4, 220–440 Hz
Overview of sound sample (25 sounds)

Female, aged 27, speaker ID 1041
Vocal range documented = 12 semitones, A_3–A_4, 220–440 Hz
Overview of sound sample (25 sounds)

Female, aged 29, speaker ID 1066
Vocal range documented = 12 semitones, A_3–A_4, 220–440 Hz
Overview of sound sample (25 sounds)

Female, aged 37, speaker ID 1071
Vocal range documented = 12 semitones, A_3–A_4, 220–440 Hz
Overview of sound sample (25 sounds)

Female, aged 34, speaker ID 1080
Vocal range documented = 12 semitones, A_3–A_4, 220–440 Hz
Overview of sound sample (25 sounds)

Female, aged 25, speaker ID 1081
Vocal range documented = 12 semitones, A_3–A_4, 220–440 Hz
Overview of sound sample (25 sounds)

7.1.2.2 Men

Male, aged 47, speaker ID 1011
Vocal range documented = 12 semitones, C_3–C_4, 131–262 Hz
Overview of sound sample (25 sounds)

Male, aged 21, speaker ID 1013
Vocal range documented = 12 semitones, C_3–C_4, 131–262 Hz
Overview of sound sample (25 sounds)

Male, aged 24, speaker ID 1015
Vocal range documented = 12 semitones, C_3–C_4, 131–262 Hz
Overview of sound sample (25 sounds)

Male, aged 27, speaker ID 1022
Vocal range documented = 12 semitones, C_3–C_4, 131–262 Hz
Overview of sound sample (25 sounds)

Male, aged 33, speaker ID 1024
Vocal range documented = 12 semitones, C_3–C_4, 131–262 Hz
Overview of sound sample (25 sounds)

Male, aged 18, speaker ID 1073
Vocal range documented = 12 semitones, C_3–C_4, 131–262 Hz
Overview of sound sample (25 sounds)

Male, aged 52, speaker ID 1075
Vocal range documented = 12 semitones, C_3–C_4, 131–262 Hz
Overview of sound sample (25 sounds)

Male, aged 35, speaker ID 1078
Vocal range documented = 12 semitones, C_3–C_4, 131–262 Hz
Overview of sound sample (25 sounds)

Male, aged 25, speaker ID 1084
Vocal range documented = 12 semitones, C_3–C_4, 131–262 Hz
Overview of sound sample (25 sounds)

Male, aged 25, speaker ID 1085
Vocal range documented = 12 semitones, C_3–C_4, 131–262 Hz
Overview of sound sample (25 sounds)

7.1.2.3 All speakers

Women, aged 20 to 37, speaker ID's =1012, 1016, 1020, 1021, 1040, 1041, 1066, 1071, 1080, 1081
Vocal range documented = 12 semitones, A_3–A_4, 220–440 Hz
Overview of sound sample (250 sounds)

Men, aged 18 to 52, speaker ID's =1011, 1013, 1015, 1022, 1024, 1073, 1075, 1078, 1084, 1085
Vocal range documented = 12 semitones, A_3–A_4, 131–262 Hz
Overview of sound sample (250 sounds)

All nonprofessional adults, aged 18 to 52, speaker ID's =1011, 1013, 1015, 1016, 1019, 1020, 1021, 1022, 1024, 1040, 1041, 1066, 1071, 1073, 1075, 1078, 1080, 1081, 1084, 1085
Vocal range documented = 12 semitones, A_3–A_4, 131–440 Hz
Overview of sound sample (500 sounds)

8 Database content IV: Summary

8.1 Main body

All nonprofessional speakers, aged 7 to 40, speaker ID's = 1009, 1027, 1034, 1036, 1037, 1038, 1039, 1044, 1045, 1051, 1054, 1056, 1057, 1058, 1063, 1088
Vocal range documented = F_2–C_6, 87–1047 Hz
Overview of sound sample (8387 sounds)

All ST speakers, aged 26 to 51, speaker ID's = 1003, 1046, 1047, 1048, 1049, 1050, 1052, 1053
Vocal range documented = C_2– C_6, 65–1047 Hz
Overview of entire sound sample (8832 sounds)

All CS speakers, aged 26 to 51, speaker ID's = 1001, 1002, 1006, 1023, 1030, 1031, 1033, 1064
Vocal range documented = D_2–C_6, 73–1047 Hz
Overview of entire sound sample (8664 sounds)

All EC speakers, aged 25 to 56, speaker ID's = 1004, 1005, 1007, 1032, 1042, 1060, 1061, 1102
Vocal range documented =D_2–C_6, 73–1047 Hz
Overview of entire sound sample (7828 sounds)

All speakers of the main body, aged 7 to 56, speaker ID's = 1001, 1002, 1003, 1004, 1005, 1006, 1007, 1009, 1023, 1027, 1030, 1031, 1032, 1033, 1034, 1036, 1037, 1038, 1039, 1042, 1044, 1045, 1046, 1047, 1048, 1049, 1050, 1051, 1052, 1053, 1054, 1056, 1057, 1058, 1060, 1063, 1088, 1064, 1102, 1103
Vocal range documented =C_2–C_6, 65–1047 Hz
Overview of entire sound sample (33711 sounds)

8.2 Side body, reference group

All speakers of the side body, aged 7 to 52, speaker ID's = 1011, 1012, 1013, 1015, 1016, 1020, 1021, 1022, 1024, 1040, 1041, 1066, 1071, 1073, 1075, 1078, 1080, 1081, 1084, 1085, 1089, 1090, 1091, 1092, 1093, 1095, 1096, 1097, 1098, 1101
Vocal range documented = 12 semitones, A_3–C_5, 131–523 Hz
Overview of sound sample (830 sounds)

9 Accessibility and terms of use

Database and recordings can be downloaded. The download consists of the database information in TXT format and the recordings in WAV format. However, the following restrictions apply:

- The use of the database is limited to scientific purposes only.
- The identity and affiliation of the users must be identifiable.
- Any publication of the results of an investigation must be in open-access form.
- Any publication of the results of an investigation must give reference to the corpus (form of citation, see title page of this website).

For a download request, please refer to the link “Terms of use” in title page of this website.

10 Acknowledgement

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

Appendix

A1 Fundamental frequencies investigated and notation of C major scale

Hz	notation	Hz	notation	Hz	notation	Hz	notation	Hz	notation
065	C_2	131	C_3	262	C_4	523	C_5	1047	C_6
073	D_2	147	D_3	294	D_4	587	D_5		
082	E_2	165	E_3	330	E_4	659	E_5		
087	F_2	175	F_3	349	F_4	698	F_5		
098	G_2	196	G_3	392	G_4	784	G_5		
110	A_2	220	A_3	440	A_4	880	A_5		
123	H_2	247	H_3	494	H_4	988	H_5		

A2 Taxonomy of professional actresses/actors and singers

For the classification of actresses/actors and singers, the below taxonomy was created based on the system of Bunch and Chapman (2000; for the full reference, see Chapter 1) but further adapted for the professional speakers investigated.

2 International

2-A International, World

2-AA ST speakers (actresses/actors)
- 2-AAA Actor/actress, major and minor principals
- 2-AAB Actor/actress, minor principals only

2-AB CS speakers (singers)
- 2-ABA Contemporary musical / musical theatre
  - 2-ABAA Musical theatre performer in a leading role
  - 2-ABAB Musical theatre performer in a supporting role
- 2-ABB Rock/pop performance and/or recording artist
- 2-ABC Jazz artist performance and/or recording artist

2-AC EC singers
- 2-ACA Opera
  - 2-ACAA Actor/actress in a leading role
  - 2-ACAB Actor/actress in a supporting role
- 2-ACB Concert/oratorio/recital principal, major principal

2-B International, Europe

2-BA ST speakers (actresses/actors)
- 2-BAA Actor/actress in a leading role
- 2-BAB Actor/actress in a supporting role

2-BB CS speakers (singers)
- 2-BBA Contemporary musical / musical theatre
  - 2-ABAA Musical theatre performer in a leading role
  - 2-ABAB Musical theatre performer in a supporting role
- 2-BBB Rock/pop performance and/or recording artist
- 2-BBC Jazz artist performance and/or recording artist

2-BC EC singers
- 2-BCA Opera
  - 2-BCAA Actor/actress in a leading role
  - 2-BCAB Actor/actress in a supporting role
- 2-BCB Concert/oratorio/recital principal

3 National/Big City

3-A ST speakers (actresses/actors)

- 3-AA Actor/actress in a leading role
- 3-AB Actor/actress in a supporting role

3-B CS speakers (singers)

- 3-BA Contemporary musical / musical theatre
  - 3-BAA Musical theatre performer in a leading role
  - 3-BAB Musical theatre performer in a supporting role
- 3-BB Rock/pop performance and/or recording artist
- 3-BC Jazz artist performance and/or recording artist

3-C EC singers

- 3-CA Opera
  - 3-CAA Actor/actress in a leading role
  - 3-CAB Actor/actress in a supporting role
- 3-CB Concert/oratorio/recital principal

T Teachers

T-A Teachers at art universities or theatre or singing academies, in parallel to stage performances

- T-AA Professor degree
- T-AB Teacher/lecturer degree

A3 Screen of the listening test

The listeners performed the listening tests remotely online (password protected), using a personal computer and headphones. The screen used for the test is shown below.
The text line line on the top of the screen indicates the listener ID on the left, the identification subseries, the number of sounds already identified ("Votes"), the number of sounds to identify and the total number of sounds of the subseries in the middle, and the actual sound (database record number) on the right.
The first line of buttons consists of the "Back", "Mark", "Play" and "Index" buttons and the "Delay" field:

- Using the "Back" button, the listener could go back to the previous sound (only), in order to allow for mistyping being corrected. However, the listener was not allowed to go back further in the substest.
- Using the "Mark" button, the listener could mark a sound in order to indicate that the investigators should double-check the sound characteristics, e.g., background noise.
- Using the "Play" button, the listener could repeat the playback the sound. No restriction was made for the playback repetition.
- When the listener clicked on a vowel button to assign a perceived vowel quality, automatically, the next sound was played back. However, after the click, a delay time was added which could be set by the listener using the "Delay" field.
- Using the "Index" button, the listener could quit the substest and go back to the subtest listing.

Using the buttons in square format, the listener could assign a single vowel quality perceived, or any combination of two neighbouring qualities.
Using the button "other (n.s.)", the listener could indicate that she or he did not recognise a vowel quality (result = "not specified").
Using the the field "Text", the listener could indicate a recognition of vowel quality other than indicated by the square buttons, e.g., an uncertain recognition of /u/ or /i/.