

The Zurich Corpus of Vowel and Voice Quality

Version 2
zhcorpus.org

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

Handbook

(Last update 2024-10-09)

Note

Part of the description of the Zurich Corpus is given in Maurer (2024)
and is referred to accordingly:

Maurer, D. (2024). *Acoustics of the Vowel – Indices*. Lausanne, Switzerland: Peter Lang Verlag. Retrieved Sep 25, 2024, from <https://www.peterlang.com/document/1462230>

INTRODUCTORY NOTE	4
1. THE ZURICH CORPUS	5
1.1 Background	5
1.2 The work version of the database	5
1.3. Published Version 1 of the Zurich Corpus	6
1.4 Published version 2 of the Zurich Corpus	7
2 TERMS AND ABBREVIATIONS	8
2.1 Entire listing	8
2.2 Speakers and styles	9
2.3 Vowel qualities and vowel notation	9
2.4 Production parameters	9
2.5 Sound status and sound ranking	10
2.6 Phoneme intended, speech intended	11
3 PART 1 OF THE ZURICH CORPUS – DETAILS OF METHOD	12
3.1 Speakers and utterances	12
3.2 Recordings	16
3.3 Acoustic analysis	20
3.4 Listening test for vowel quality recognition	22
3.5 Sounds recorded and sounds selected for publication	23

4 PART 1 OF THE ZURICH CORPUS – LIST OF SPEAKER-RELATED SOUND SAMPLES	25
4.1 Nonprofessional speakers of the main body of part 1	25
4.1.1 Children	25
4.1.2 Adult speakers	25
4.1.3 All nonprofessional speakers of the main body of part 1	26
4.2 Professional speakers of the main body of part 1	27
4.2.1 ST speakers	27
4.2.2 CS speakers	29
4.2.3 EC speakers	31
4.3 Nonprofessional speakers of the side body of part 1 (reference group)	33
4.3.1 Children	33
4.3.2 Adults	34
4.4 Summary	36
4.4.1 Main body	36
4.4.2 Side body	36
5 PARTS 2 TO 5 OF THE ZURICH CORPUS – DETAILS OF METHOD	37
AND LINKS TO SOUND SAMPLES	37
6 SOFTWARE TOOLS IMPLEMENTED INTO THE CORPUS	38
7 ACCESSIBILITY AND TERMS OF USE	38
APPENDIX	39
A1 Fundamental frequencies investigated and notation of musical	39
C major scale	39
A2 Taxonomy of professional actresses/actors and singers	39
A3 Screen of the listening tests	41

Introductory note

Overview

Chapter 1 provides a general description of the Zurich Corpus of Vowel and Voice Quality: the work database and its five parts of investigation, and the extracts of the first and second published version. (Note that Chapter 1.1 refers to Maurer et al., 2018, and Chapters 1.2–1.4 refers to Maurer, 2024, in terms of text extracts and their adaptation to this handbook.)

All the following chapters concern the second published version online.

Chapter 2 lists and details the terms and abbreviations used here.

Chapters 3 and 4 concern the first part of the corpus. For this part, Chapter 3 details speakers, utterances, recordings, acoustic analysis, listening test and sound selection for publication, and Chapter 4 lists sound samples related to speaker groups and single speakers and provides corresponding links.

Chapter 5 details speakers and utterances of the second to the fifth part of the corpus.

Chapter 6 describes the software tools integrated into the corpus. (Note that the text of this chapter refers to Maurer, 2024.)

Chapter 7 provides information about the accessibility of the corpus and the terms of use.

The Appendix provides details of (i) fundamental frequencies (hereafter f_0) investigated and notation of musical C major scale, (ii) the taxonomy of professional actresses/actors and singers applied, and (iii) the screen used for the standard listening test performed.

Main references (in addition to Maurer, 2024)

Maurer, D., d’Heureuse, C., Suter, H., Dellwo, V., Friedrichs, D., & Kathiresan, T. (2018). The Zurich Corpus of Vowel and Voice Quality, Version 1.0. In *Proceedings of Interspeech 2018* (pp. 1417–1421). Retrieved Sep 25, 2024, from <https://zhcorpus.org/v2/doc/MaurerEtAl2018.pdf>

Maurer, D. (2016). *Acoustics of the Vowel*. Lausanne, Switzerland: Peter Lang Verlag. Retrieved Sep 25, 2024, from <https://www.peterlang.com/document/1067698>.

Acknowledgement

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

1. The Zurich Corpus

1.1 Background

“Besides a great many databases of continuous speech, numerous samples or 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.“ [...] However, in general, [these] existing samples or databases either present sounds produced by speakers with a medium vocal effort at particular f_0 , or they compare sounds of speakers related to only two different production parameters, e.g., voiced and whispered phonation, or V and CVC context, or voiced with varying vocal effort, or voiced with varying f_0 in singing, etc. (for references, see Maurer et al., 2018). 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, [hitherto, we did] 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, and that can serve as an empirical reference for future research and new approaches.” (Maurer et al., 2018; see this paper for further details, examples and references.)

1.2 The work version of the database

Against this background, the Zurich Corpus of Vowel and Voice Quality – in short, the Zurich Corpus – was created in the form of an extensive unpublished sound database (hereafter work version), with selected, smaller versions thereof published online with open access. The database is still being extended continuously. The online version 1 was already published earlier (see Maurer et al., 2018). The online version 2 was published in the context of a treatise entitled Acoustics of the Vowel – Indices (see Maurer, 2024).

The entire sound database (work version) consists of five different parts:

- Part 1 Natural vowel sounds, produced by single speakers with a systematic variation of basic production parameters; in addition, for each of the speakers, a read reference text (“Nordwind und Sonne”, see Handbook of the International Phonetic Association, 1999, pp. 88–89) and one or several songs sung were also recorded
- Part 2 Extracts of speech and singing documented from everyday utterances and utterances in the field of the performing arts
- Part 3 Syllables and minimal pairs produced by single speakers at different f_0 levels
- Part 4 Manipulated natural, resynthesised and synthesised sounds
- Part 5 Miscellaneous

The first part addresses the question of observable acoustic characteristics of vowel sounds. The second part addresses the question of the observable f_0 ranges in intelligible speech and singing. The third part documents vowel sounds produced in the specific context of syllables and minimal pairs by single speakers at various f_0 levels. The fourth part documents sounds investigated in the context of different experiments related to sound filtering, resynthesis and synthesis. The fifth part consists of miscellaneous sounds that were cast aside during the creation of the corpus.

Only the first part is both extensive and systematic in its structure. The second part aims at a documentation of speech and occurring f_0 ranges and at highlighting the significance of

extensive f_0 variation. However, this documentation is not of a systematic structure. (Future language-specific databases will have to address the question of a systematic sample of utterances that could serve as an empirical reference for observable f_0 variation in intelligible speech.) The remaining three parts document experiment-specific sounds and sounds that were recorded with limited variation of production parameters or in an unsystematic way.

Details of all five parts of the sound database (working version, unpublished, including all recordings made up to 2022) are given in Maurer (2024, pp. 36-41).

Details on the acoustic analysis and the standard listening test conducted are given below (see Chapters 3.3 and 3.4).

A Note on the redundancy of recorded sounds:

In numerous cases, we kept duplicate recordings for one single production task (above all for the sounds in the first, third and fifth part): If either the speaker or the investigator believed that a repeated recording of a specific utterance could improve sound and/or vowel quality, the recording was repeated one or several more times. For further details and the sound selection for publication, see the next chapter and Chapter 3.5.

1.3. Published Version 1 of the Zurich Corpus

Details of the first published version of the Zurich Corpus are given in Maurer (2024, pp. 42-43).

A note on the sound selection for Part 1: As mentioned, “[...] in many cases, two or multiple recordings were made for a single speaker and a specific configuration of production parameters when creating the database, a systematic subset of sounds was compiled for publication: If only one sound was recorded for a specific configuration of production parameters, then this sound was selected; if two or more recordings for a specific configuration of production parameters were made, then the sound with the highest recognition rate, the longest duration and the smallest difference between the f_0 intended by the speaker and the average f_0 calculated for the radiated sound was selected (according to this order). For nonstyle productions and each level of vocal effort separately, the sound selection was further limited to an f_0 range for which, for a single speaker, all vowels investigated were represented by a sound. (Note that not all speakers were able to produce a complete set of investigated vowels at the very lowest or highest limit of their vocal range. Therefore, f_0 levels with incomplete vowel sounds for a specific production parameter setting were excluded.) For productions in style mode, the f_0 range was generally set at the discretion and the style-specific range of the artist.” (Maurer, 2024, p. 42)

For the sound selection procedure, see also Chapter 3.5.

As a result, a systematic sample with one sound per production task for the investigated f_0 range of a speaker (if available for all eight long Standard German vowels) was created for Part 1 of the published corpus.

1.4 Published version 2 of the Zurich Corpus

Details of the first published version of the Zurich Corpus (related to the date of publication) are given in Maurer (2024, pp. 43-46), including three tables that provide an overview of the general structure of the corpus.

A note on playback functionality: Note that open-access playback functionality is given only for the sounds for which, in the figures and the tables, sound links are provided. For all other sounds, a login is required (see the corresponding menu item). Without login, the sound player is disabled and, as a placeholder, a link to the Terms of Use of the corpus is provided. This document lists the conditions for general open access to the playback functionality for all sounds of the corpus and informs about the login request procedure (see also the title page of the Zurich Corpus, menu item Terms of Use). Furthermore, the document lists the conditions for downloading the entire database for scientific use.

2 Terms and abbreviations

2.1 Entire listing

Below, all abbreviations used here are listed in alphabetical order. Details are provided in the subsequent chapters.

ar	artificial (synthesis, different synthesis methods)
A	Adults (age-related speaker group)
b	breathy phonation
c	creaky phonation (vocal fry)
C	Children (age-related speaker group)
CV	vowel sound produced in CV context (vowel sound context, consonant–vowel)
CO	Cantonese opera singing/speaking style
CS	contemporary singing style (generic term, including the substyles contemporary musical theatre, pop, rock, and jazz)
D	standard recording, but ‘deficient’ (minor sound impairment, flagged accordingly)
EC	European classical singing style
F1, F2, ...	formant frequencies
fo	fundamental frequency
hgh	high vocal effort
hp	HP filtered.
i	isolated (extracted sound parts of speech)
K	recordings made in a concert hall
kl	Klatt synthesis
low	low vocal effort
lp	LP filtered
m	man or men for male speaker(s) (indication of gender)
m	mixed (indication of phonation)
M	miscellaneous (recordings made under various conditions)
med	medium vocal effort
mp	vowel sound produced in minimal pair context (vowel sound context)
ms	intended f_0 according to the musical C-major scale
N	nonstyle (indication of production style)
N	nonprofessional (indication of a speaker subgroup)
nc	extracted sound nucleus
ns	not specified
low	low vocal effort
r(i)	ranking order
rp	intended f_0 according to a reference pitch given
S	standard recording
sh	shouted
ST	straight theatre speaking style
sVsV	vowel sound produced in s-V-s-V context (vowel sound context; s = consonant /s/, V = /vowel/)
T	test sounds (recordings made/sounds produced for specific investigations)
text	vowel sound produced in text context (indication of the context = speech or singing)
tr	text read (indication of the production mode)
ts	text sung (indication of the production mode)
v	voiced phonation
V	in vowel context (single vowel sound including on- and offsets)
VI	voice imitation (utterances of impressionists)
var	varied vocal effort
w	woman or women for female speaker(s) (indication of gender)
w	whispered phonation

2.2 Speakers and styles

The database presented includes systematic 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, rock, and jazz), and singers of European classical singing (EC).

Production styles are notated correspondingly: nonstyle (N), straight theatre speaking style (ST), contemporary singing style (CS), European classical singing style (EC).

The database also includes unsystematic recordings of utterances that are produced in other styles. Two of these styles are indicated: Cantonese Opera style (CO) and Voice Imitation (VI, utterances of impressionists). Other styles are either not further specified (ns) or are commented in the sound information.

In this handbook, both speakers and singers are referred to as speakers as a generic term.

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

2.3 Vowel qualities and vowel notation

The main focus of the database concerns vowel qualities of the long Standard German vowels /i–y–e–ø–ɛ–a–o–u/. Included is the vowel /a/, which can be encountered as a long vowel in some regions of Germany, Austria and Switzerland and also in some singing styles. Therefore, the indication /a/ in this text refers to the vowel area /a–ɑ/, that is, including all allophones of /a/ or /ɑ/.

In the database, the above vowel qualities are notated as /i–ü–e–ö–ä–a–o–u/, that is, using German characters.

A few additional sounds concern the two vowels /ə/ and /ɔ/, notated as /e1/ and /o1/.

2.4 Production parameters

The database presents recordings of utterances produced with varying basic production parameters such as phonation type, production mode, vowel context, vocal effort, fundamental frequency and speaking or singing style.

Phonation types and related abbreviations:

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

Production modes and related abbreviations:

tr	text read
ts	text sung
rp	f_0 according to a reference pitch given
ms	f_0 according to musical scale (C-Major)
sh	shouted
ns	not specified (no specific production mode)
ar	artificial (resynthesis, synthesis, different synthesis methods)
hp	HP filtered sounds
lp	LP filtered sounds

Vowel sound context and related abbreviations:

V	in vowel context (single vowel sound including on- and offsets)
sVsV	in s-V-s-V context (s = consonant /s/, V = /vowel/)
mp	in minimal pair context
nc	extracted sound nucleus
text	in read/sung text or free speech context

Vocal effort and related abbreviations:

med	medium
low	low
hgh	high
var	varying
ns	not specified

Intended fundamental frequency f_0 is indicated according to the musical C-major scale (see Appendix), and calculated f_0 is indicated according to the result of the standard acoustic analysis, as detailed in Chapter 3.3.

For the production styles and related abbreviations, see Chapter 2.2.

2.5 Sound status and sound ranking

The sounds presented in the database are classified according to the condition of their recording or their production (resynthesis, synthesis, sound filtering), termed here sound status. The corresponding abbreviations are used for the different parameters:

Sound status and related abbreviations:

S	standard recordings (recorded under standard conditions and with standard production parameters)
D	standard recordings, but manifesting minor impairments
M	miscellaneous (recordings made under various conditions)
K	recordings made in a concert hall
T	test sounds (recordings made/sounds produced for specific investigations)

In addition, a sound ranking was conducted for the selection of the sounds to be published. The corresponding procedure and the related abbreviations are given in Chapter 3.5.

2.6 Phoneme intended, speech intended

//Text// text read or text sung (indication for intended speech)

3 Part 1 of the Zurich Corpus – Details of method

In this main chapter, details of method are described concerning the sounds of part 1 of the corpus as introduced in Chapter 1.3. Note again that parts of the text are taken from Maurer (2024, Chapter 1.1).

3.1 Speakers and utterances

Overview

For the general content of the Zurich Corpus and the systematic sound compilation of its first part, see Maurer (2024, pp. 45–46). These two pages are inserted below.

Table 1. The Zurich Corpus, Version 2: General content. Column 1 = parts of the corpus. Column 2 = content of the parts. Columns 3 and 4 = links to the sounds and the speakers.

[C-01-01-T01]

Table 2. The Zurich Corpus, Version 2, Part 1: Speakers. Column 1 = style-related speaker groups (nonprofessionals and professionals; ST = Straight Theatre, CS = Contemporary Singing styles, EC = European Classical singing style). Columns 2–6 = number and age range of speakers (children and adults; f = female speakers, m = male speakers) of the main body (vowel sounds produced with extensive variation of production parameters) and the side body (voiced vowel sounds produced with limited f_0 variation).

[C-01-01-T02]

Table 3. The Zurich Corpus, Version 2, Part 1: Production parameters. Column 1 = phonation type. Column 2 = vocal effort. Column 3 = vowel context (V = in isolation, sVsV = in /s/–V–/s/–V context). Column 4 = intended f_0 level variation (musical scale = according to C-major scale; upper scale and reference f_0 , see text). Column 5 = production style (N = nonstyle, S = style of ST, CS or EC).

[C-01-01-T03]

Links to the Zurich Corpus, Version 2

Title page: [↗](#)

Assistant (Help menu): [↗](#)

How to search (Help menu): [↗](#)

Lightbox help (Help menu): [↗](#)

Abbreviations: [↗](#)

Handbook: [↗](#)

Table 1. The Zurich Corpus, Version 2: General content. [C01-01-T01]

Part	Content	Sounds	Speakers
Part 1	Vowel sounds, text read, songs sung, systematic recordings	↗	↗
(thereof)	One sound per production parameter configuration (revision of version 1 of the Zurich Corpus; details see Tables 2 and 3)	↗	↗
(thereof)	Additional sounds related to this treatise and the Additions of the Zurich Corpus	↗	↗
Part 2	Speech extracts	↗	↗
Part 3	Minimal pairs	↗	↗
Part 4	LP/HP filtered, resynthesised and synthesised sounds	↗	↗
Part 5	Miscellaneous	↗	↗
Entire Corpus (Maurer et al., 2024)		↗	

Table 2. The Zurich Corpus, Version 2, Part 1: Speakers. [C01-01-T02]

Speaker group	Children			Adults			Speakers (total)
	f	m	age	f	m	age	
Main body							
Nonprofessionals	4	4	7–10	4	4	23–40	16
ST actresses/actors	–	–	–	4	4	26–51	8
CS singers	–	–	–	4	4	26–50	8
EC singers	–	–	–	4	4	25–56	8
Side body							
Nonprofessionals	5	5	7–9	10	10	18–52	30
Total	9	9	–	26	26	–	70

Table 3. The Zurich Corpus, Version 2, Part 1: Production parameters. [C01-01-T03]

Phonation	Vocal effort	Vowel context	fo intended	Production style
voiced	medium	V	musical scale	N / S
voiced	low	V	musical scale	N / S
voiced	high	V	musical scale	N / S
voiced	medium	sVsV	upper scale	N / S
voiced	shouted	V	–	N
voiced	shouted	sVsV	–	N
breathy	low	V	–	N
creaky	medium	V	–	N
whispered	medium	V	–	N
whispered	medium	sVsV	–	N
voiced	medium	V	reference fo	N

Speakers and utterances of extensive investigation (main body, see the above Tables 2 and 3)

For the speakers of extensive investigation, see Maurer (2024, pp. 36–37). All adult speakers were remunerated with a participation fee. The children obtained a small gift.

Reference speakers and utterances (side body)

For the reference speakers and utterances, see Maurer (2024, pp. 37–38).

Further details

Acquisition and selection of speakers: The acquisition and selection of speakers proved to be a highly sensitive topic. Various conditions had to be met by potential speakers: For the professional singer's group (main body, adults only), our requirements were as follows: highly trained voice and proof of professional activity on stage as a performer/singer (predominantly in German language), ideally speakers with German as their first language (or if not, good command of German phonetic articulation on stage; see below for 'professional speakers with first languages other than German'), vocal range of at least 2 octaves, openness and necessary skill set to explore the production of nonstyle utterances (technical and physiological versatility of voice apparatus and singing style), geographic availability and willingness to participate for the fee provided.

For the nonprofessional speaker's group (main body, adults and children), our requirements were as follows: speakers with German as their first language (including Swiss German), vocal range of at least 2 octaves (adults) and at least 1.5 octaves (children), good musical ear, ability to imitate a vowel quality auditioned by an investigators vowel sound, ability to sing/speak on a pitch presented either orally by the investigator or by using a digital piano, openness and necessary skill set to uninhibitedly explore vowel production on uncommon pitches (very low, very high), geographic availability and willingness to participate for the fee provided (adults) or for free (children).

For the reference speaker's group (side body, adults and children), our requirements were as follows: speakers with Swiss German dialects from the eastern Swiss midland as their first language, ability to produce recognisable vowel sounds on a specific pitch over an f_0 range of at least 15 semitones, ability to imitate a vowel quality auditioned by an investigators vowel sound, ability to sing/speak on a pitch presented either orally by the investigator or by using a digital piano, geographic availability as well as willingness to participate for free.

Professional speakers with first languages other than German: As mentioned, four professional singers are not speakers with German as their first language but they have a professional track record of performing live on stage in Standard German. Their corresponding pronunciation ability can be evaluated on the bases of the read and sung texts. Further, female singer 1005 teaches singing in an Art University in Switzerland, and female singer 1031 is a well known performer in German musical theatre plays.

Vowel qualities: The phonetic (and perceptual) distance between /o/ and /a/ is very pronounced. Therefore, in the initial methodological design, we decided to elicit utterances of the vowel /ɔ/ as a long, sustained vowel. However, since in the German language, /ɔ/ is only used as short vowel, some speakers exhibited difficulties to maintain a constant vowel quality for /ɔ/. In

consequence, we decided not to include these sounds into the systematic part of the corpus. However, in the listening test, the vowel was consistently included to test vowel boundaries /o–ɔ/ and /ɔ–a/ (see below).

Vocal effort variation: Speakers were asked to produce sounds with a comfortable medium vocal effort and medium loudness, with a low vocal effort and soft loudness, and with a high vocal effort and high loudness while avoiding any harm the vocal cords. No objective reference was given to the speakers, and their reaction accorded to their understanding of the task and their ability of vocal effort variation for vowel sounds of different qualities and at different f_0 .

Shouted vowel sounds: For shouting, each speaker was asked to spontaneously shout vowel sounds. In consequence, there was no predefined level of f_0 , and the actual f_0 levels for the eight vowel sounds of a series corresponds only approximately. However, in the editing process, an average “intended” f_0 level was assigned to the sounds of one series.

If a speaker spontaneously proposed substantially different levels of f_0 for shouting, two or three sound series for the eight vowels were recorded on two or three substantially different f_0 levels.

Note that due to this procedure, the f_0 levels of shouted vowels are only specific to a certain sound set of a single speaker and not to all speaker groups.

Vowel sounds produced with creaky phonation: For nonprofessional speakers (main body), vowel sounds with creaky phonation were not included in the standard recordings of investigation. However, during the recordings, a few nonprofessional speakers demonstrated the ability to produce creaky sounds, in which case we decided to record said sounds and document them in the corpus.

Range of f_0 for specific artistic speaking or singing styles: As mentioned, for the sounds produced in a professional style of a singer or actress or actor, the f_0 range was set as follows: For ST and CS styles, the vocal range was set according to the individual, professional judgment of the speaker; for EC, the ranges were set to G2–G4 (98–392 Hz) for baritones, C3–C5 (131–523 Hz) for tenors, A3–A5 (220–880 Hz) for mezzo-sopranos, and C4–C6 (262–1047 Hz) for sopranos.

3.2 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_0 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 a similar manner. In order to subsequently determine the actual sound pressure level, for each recording session, a 1 kHz pure tone was recorded with a reference gain using a sound level calibrator (Brüel & Kjaer 4230).

Recording procedure

Utterances were recorded according to a specific configuration of production parameters (see Table 3), separating nonstyle and style productions.

For sounds in V context, except for shouting, the speakers were asked to monotonously sustain a sound on a given f_0 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 further details, see below.)

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 (see above, the note on the redundancy of recorded sounds).

f_0 scale

For each speaker, a comfortable “middle” pitch on the C-major scale was determined from which vocalisations were then produced up and down this 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_0 autonomously. If not, each f_0 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 standard listening test (see below) was performed for each speaker. If the listening test did not yield satisfactory results (low recognition rate) or the sound duration was very short, 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 nonstyle 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 sounds were extracted from the recorded sound series using a semi-automatic tool (proprietary development) and procedure: In a first step, each single utterance was saved in a separate WAV file. Subsequently, these WAV files were visually crosschecked and further edited so as to successfully retain on- and offsets and, for V context, to approximately center the vowel sound as a basis for automatic sound nucleus determination (see below, acoustic analysis).

Editing problems occurred if the on- or offset of a sound was tainted with room or speaker noise (e.g., reverberation or breathing, blowing). In this case, the noise was often included in the sound file.

Some speakers produced small slips of the tongue while reading the text. In these cases, during the recording session, the speakers were asked to do a retake of the text passage that had not been rendered correctly. In the editing process, the slips of the tongue were then deleted and replaced with the correct retake.

Finally, each single sound file was then labelled with a database reference number and relevant sound and speaker information in anonymous form.

Further details

Recording rooms and reverberation: For the present study, anechoic chambers are unsuitable for different reasons. (i) Speakers of all speaker groups of the main body are generally rather sensitive to the absence of room acoustics that acoustically support vocal production: They react negatively to a “deaf” or sound-proof acoustic atmosphere. (ii) Recording sessions with a duration of two to four hours (including breaks), as were often the case in the present investigation, need corresponding room comfort and direct interaction between investigator and speaker. (iii) Tasks to produce voiced sounds in artistic styles and, thereby, vary basic production parameters, relate to room acoustics, and some reverberation – even limited so as not to substantially affect the sound quality for vowel and voice related acoustic analysis – is therefore needed. The same holds true for tasks to focus on the maintenance of vowel quality when varying basic production parameters. (iv) Because of number and duration of recording sessions, we were not able to book the same recording room for each session. Additionally, for some recording sessions, we were required to travel to a current location of a speaker (above all for professional singers, as their professional activity/contracts did not allow them to travel to us). As a consequence, numerous different recording rooms were used for sound recordings. However, careful attention was given to always selecting maximally quiet rooms, and – whenever possible – booking recording studios at partnering Universities or renting professional recording studios. With rare exceptions, the rooms had a weak reverberation and very little to no background noise. Nevertheless, room reverberation is sometimes perceivable, above all for sounds produced with high vocal effort or sounds recorded with a high gain setting. Such reverberation was considered in the editing process (see above). For recordings with high gain setting and very low vocal effort, background noise is sometimes perceivable before or after the actual sound.

Duration of recording sessions: For the nonprofessional speakers (main body), the standard recording session lasted approximately 3 hours, including one or two breaks. Subsequent session(s) for corrections generally lasted about 1 hour. – For professional speakers (main body), the first two standard recording sessions (two different days, one style and one nonstyle

session) lasted approximately 3 hours, including one or two breaks. Subsequent session(s) for corrections generally lasted 2 hours maximum, including a break. – For the reference speakers (side body), the duration of the recording sessions did not exceed 30 minutes.

Gain levels and sound amplitudes: Gain levels were held constant for sounds sets with increasing or decreasing f_0 . However, vocal effort often changed substantially with increasing or decreasing f_0 (i.e. louder for high sounds, softer for low sounds) and, therefore, the sound amplitudes within a sound set recorded sometimes varies strongly. – Some whispered or breathy sounds have very low sound amplitude although the gain level was set very high.

Sound duration: Ideally, individual sound duration was aimed to last at least 1 to 2 sec (including on-/offset) for sounds in V context (except for shouting) and at least 0.5 to 1 sec (including on-/offset) for sounds in sVsV context. However, intra-speaker sound duration varied heavily according to style, f_0 , vocal effort, production mode and phonation type, even for sounds produced by professional speakers.

In parallel, very pronounced inter-speaker differences in sound duration also occurred. Despite clear recording instructions, some nonprofessional speakers (adults as well as children) habitually produced rather short vowel sounds (sometimes below 1 sec for V condition and below 0.5 sec for sVsV condition), while some of the professional speakers predominantly produced long sounds (sometimes up to c. 4 sec).

In general, sounds in V condition were sustained longer than sounds in sVsV condition, and shouted sounds – and for some speakers also whispered sounds – were sometimes very short.

Against this background, no strict control of sound duration was carried out during the recording sessions because the idiosyncratic self-perception of duration proved to be highly varied and inconstant, and a close monitoring of duration while being focused on successful production of other basic production parameters proved to be too difficult. (Various attempts to control for sound duration in the initial stages of a session were made by the investigators, and at the beginning of every session, every speaker was instructed to sustain each vowel sound for 2 seconds at minimum. But if the utterances continued to remain either too short or too long, ultimately, the vocal production habits of a speaker were not compromised any further.)

In the published corpus, as mentioned above, the sound duration of vowel nuclei (excluding on- and offset) generally is ≥ 0.5 sec for V condition (with few exceptions for shouted vowels), and ≥ 0.3 sec (excluding on- and offset) for sVsV condition.

Intended f_0 according to the musical C-major scale and actual f_0 of the produced sounds: Some speakers managed to stay very true to the requested f_0 levels when singing up or down the C-major scale, while other speakers tended to go increasingly flat or sharp as the scale progressed. Because of this, and because certain sounds of a scale from an initial recording were later substituted by revised and corrected sounds, the actual series of f_0 levels of a sound series may in some cases exhibit small irregularities when compared to C-major scale.

3.3 Acoustic analysis

Analysis

For vowel sounds produced in V context, the analysis was conducted on the approximately middle 0.3 sec of a sound for a frequency range of 0–5.5 kHz on f_0 contour, average f_0 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 approximately middle 0.3 sec of the first or the second vowel sound, depending on their duration and sound quality. Concerning the allocation of the 0.3. sec nucleus of an sVsV utterance, in a first step and using a proprietary program, the first vowel nucleus in the utterance was assigned for acoustic analysis if vowel sound duration was ≥ 0.5 sec or, if not, its duration was longer than the duration of the second vowel sound and the perceived vowel quality was equal to or better than the quality of the second vowel sound. If this does not apply, the second vowel nucleus was assigned. Subsequently, the corresponding recording was visually crosschecked and, if necessary, the nucleus was manually set by the investigator (first author). – For whispered sounds, the nuclei were manually set on the basis of the duration of the vowel nuclei and their perceptual quality.

The read texts and the songs/arias were analysed for f_0 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 (Boersma and Weenink, 2020; versions used from 2014 onwards).

Reference:

Boersma, P., and Weenink, D. (2020). Praat: doing phonetics by computer [Computer program]. Version 6.1.12, retrieved 30 April 2020 from <http://www.praat.org>.

Graphic representations, numerical indications

For vowel sounds, graphic representation includes the display of the entire sound wave, the sound nucleus, the f_0 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 matched onto the spectrum in order to approximately visualise the correspondence between spectral peaks and calculated formants. Numerical average values of f_0 and formant patterns (sound nucleus analysed) were added to the sound information.

For texts and songs/arias, graphic representation included the display of the sound wave, the f_0 contour, the spectrogram and the LTAS.

Crosschecks of sound quality and acoustic analysis

All sounds were acoustically crosschecked and sounds with marked background noise unrelated to vowel sound production were removed.

Some sounds contain minor noise caused by vowel sound production, above all puffs of air (insufficiently blocked by the pop filter) or byproducts of articulation (e.g., tongue-clicks). Such noise mostly occurs in on- or offsets. They are considered as an integral part of vocal production and, therefore, are included in the sound documentation.

Some sounds contain other forms of noise that are unrelated to voice production (e.g., speaker's body movements picked up by the microphone as is often the case for children, investigators taking notes on keyboard, etc.). If these events were considered to impair vowel recognition, they were removed. If these events were very soft but were considered to potentially affect the results of the acoustic analysis related to vowel or voice quality, the sound file was flagged with the status label "D" (deficient, see below). (Note that this concerns only a few sounds in the published version of the corpus.)

Very subtle and hardly perceivable noise events, considered to not impair vowel and voice perception or vowel- and voice-related spectral characteristics, are not flagged.

Other qualitative sound aspects concern perceivable reverberation and room-related low background noise for sound recording with high gain level and low vocal effort (see above).

Numerical values were crosschecked for accuracy of average f_0 . In cases of f_0 calculation errors (above all doubling or halving of calculated f_0 when compared with intended f_0 and recognised pitch), calculated f_0 levels were manually corrected by ear, generally allocated to the nearest match of levels of the C-major scale and, in a few cases, allocated to a frequency in Hz in between semitones of the C-major scale. Note that in cases of f_0 calculation errors, the manually corrected levels are indicated in the sound information. However, in the graphic representation of automatic analysis, the erroneous f_0 contour of the automatic analysis is given.

Display of graphics and numerical indications

For details concerning the display of sound information, results of acoustic analysis and corresponding illustrations, please refer to the Assistant in the Help menu of the database.

Note that, in the graphic representations of the sounds, the amplitudes of the sounds are normalised.

Sound status

All sounds of the first part of the corpus, which were recorded according to the standard parameter setting and standard procedure, are assigned with the sound status "S" for standard.

However, as mentioned above, a few of these sounds manifested noise unrelated to vowel or voice quality. If they could not be replaced by sounds with equal vowel recognition rate, they are flagged with the status "D" (standard recording but deficient sound quality).

Additional flags are "K" ("Konzerthalle; sounds recorded in a concert hall), "T" (test sounds) and "M" (miscellaneous; various additional sounds).

3.4 Listening test for vowel quality recognition

Listeners

Phonetic expert listeners: A group of five phonetic expert listeners (professionally trained singers or actors or voice teachers) performed listening tests for the vowel sounds of the first part of the corpus. Each listener passed a pure-tone hearing screening (25 dB at octave frequencies from 0.5–4 kHz, using a Beltone 110 audiometer) to exclude hearing impairment.

Because of the high number of sounds subjected to listening tests (recognition of vowel quality), the extensive variation of production parameters, the testing of vowel boundaries and the long period of data collection, only a small number of five listeners, all professionally trained as speakers or singers, performed the vowel quality recognition test.

Constancy of listeners: Because of the long period of data collection, sound editing and continuous listening tests, three of the five listeners participated in all subtests, one listener was replaced once, and another listener was replaced three times, to keep the number of listeners constant; however, all listeners were professionally trained speakers or singers, as described above. Throughout the course of investigation, the gender distribution among the listener group was either three women and two men or two women and three men. The first and third author were part of the listener group.

Screen used by the listeners: Appendix A3 gives an overview of the type of web-application and visualisation tool for the listening tests.

Test procedure, vowel recognition rate

Testing vowel recognition was organised into speaker-specific subtests (blocked-speaker condition), further separating nonstyle 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 judge each of the vowel sounds presented and to assign them to one of the following categories: (1) a single specific Standard German vowel (/i–y–e–ø–ɛ–a–ɔ–o–u/), (2) a vowel boundary region of two vowels, (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 assignment of the vowel /a/ included all variants in the region of /a–ɑ/ because the production of this vowel varies strongly among German speakers.

The vowel /ɔ/ was included in the listening test because the perceptual distance /a–ɔ/ is very large, not representing comparable vowel quality distance comparable to /o–u/, /i–y/, /i–e/, /y–ø/, /e–ø/, /e–ɛ/, /ø–ɛ/, /ɛ–a/.

The term vowel recognition rate is used for the ratio or percentage of listeners who assigned a certain vowel quality. In most cases, the rate is given as a percentage even if the rate relates to the recognition results of the five standard listeners. For the results of the standard listening

test, a vowel quality is said to be recognised if 3 (or more) of the 5 listeners assigned the same vowel quality, and the recognition rate is given in % (60% = 3 of 5 listeners, 80% = 4 of 5 listeners, 100% = 5 of 5 listeners).

However, in the corpus (see the sound information and the Breakdown submenu item in the Statistics menu item), the vowel recognition rate refers to the intended vowel quality, and the recognition rate is given in % according to vowel intention.

3.5 Sounds recorded and sounds selected for publication

For the 70 speakers and systematic recordings of the first part of this second version of the corpus, in total, c. 56'600 recordings were made (excluding the recordings not matching the documented production parameters). 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 according to a ranking of the sounds. In short, as indicated above, if only one sound is documented in the work database for a specific configuration of production parameters, then this sound was selected; if there are multiple recordings for one specific configuration of production parameters, 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 more details, see below.

For nonstyle productions and each vocal effort separately, the sound selection was further limited to an f_0 range in which all vowels investigated were represented. (For example, if at a very low or high f_0 , a sound of /i/ could not be produced, then this level of f_0 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 limited to corresponding style-specific f_0 ranges as practiced by the artist in question (see Chapter 3.1).

Additional information concerning the sound selection (ranking)

Concerning the sound ranking applied for sound selection, each ranking category relates to a certain configuration of production parameters, represented in a single position in the production matrix; the ranking thus concerns either a single sound or several sounds recorded for a specific configuration of production parameters. The aim of the ranking is to create an objective procedure for selecting the 'best' sound (the sound that best corresponds to a phenomenological investigation) for each speaker and each production parameter for the publication of the corpus and to qualify the vowel recognition for the selected sound.

The applied ranking categories and their abbreviations used in the corpus are as follows:

- r0 = sounds that have not undergone a listening test; this ranking concerns all sounds of read text and songs.
- r1, r2, r3, r4 = one sound or several sounds is/are documented in the corpus for a given configuration of production parameters and, thereof, one or several sounds has/have a recognition rate $\geq 60\%$.
 - r1 = the sound with the highest vowel recognition rate $\geq 60\%$, the longest duration and the smallest difference of f_0 intended and f_0 calculated is selected and assigned r1 status.
 - r2 = the sound with the second highest vowel recognition rate $\geq 60\%$ or the second longest duration or the second smallest difference of f_0 intended and f_0 calculated is assigned r2 status.
 - r3 = the sound with the third highest vowel recognition rate or the third longest duration or the third smallest difference of f_0 intended and f_0 calculated is assigned r3 status.
 - r4 = the sound(s) with a vowel recognition rate $< 60\%$ is/are assigned r4 status.
- r5, r6, r7 = one sound or several sounds is/are documented in the corpus for a given configuration of production parameters but, thereof, none has a recognition rate $\geq 60\%$.
 - r5 = if only one such sound is documented in the corpus, it is assigned r5 status.
 - r6 = if two or more sounds are documented in the corpus, one of these sounds is manually selected by the author and is assigned r6 status; manual selection is made with regard to the details of vowel recognition, difference of f_0 intended and f_0 calculated, sound duration and sound quality.
 - r7 = other sounds with vowel recognition $< 60\%$.

For the first part of the published corpus, sounds with the ranking status r0, r1, r5 and r6 were selected in order to provide a systematic sound configuration. Thus, for each speaker, each style and each single production parameter configuration, one sound file is presented.

According to this selection, for a specific configuration of production parameters, that is, a specific position in the production matrix, the rankings indicate:

- r0 = texts read and sung
- r1 = sound with the highest vowel recognition rate, the longest duration and the best match of f_0 intended and f_0 calculated (in this order)
- r5 = selection of the single sound existing for a position, with no vowel recognition according to vowel intention
- r6 = selection of one of several sounds related to a position, with no vowel recognition for all of these sounds

For the other parts of the corpus, the ranking status plays no role for the sound selection.


4 Part 1 of the Zurich Corpus – List of speaker-related sound samples

4.1 Nonprofessional speakers of the main body of part 1

4.1.1 Children


Female, aged 8, speaker ID 1009

Vocal range documented = G₃–A₅, 196–880 Hz

 [Overview of sound sample \(539 sounds\)](#)

Female, aged 10, speaker ID 1034


Vocal range documented = F₃–A₅, 175–880 Hz

 [Overview of sound sample \(498 sounds\)](#)

Note: Some sounds contain reverberation.


Female, aged 8, speaker ID 1037

Vocal range documented = G₃–A₅, 196–880 Hz

 [Overview of sound sample \(514 sounds\)](#)

Female, aged 10, speaker ID 1038


Vocal range documented = G₃–G₅, 196–784 Hz

 [Overview of sound sample \(420 sounds\)](#)

Note: Sounds with marked reverberation; numerous sounds with short duration.

Male, aged 8, speaker ID 1054


Vocal range documented = G₃–F₅, 196–698 Hz

 [Overview of sound sample \(434 sounds\)](#)

Note: Some sounds contain reverberation; whispered sounds contain an electronic hum (apply HP filtering with cutoff = 120 Hz to clear)


Male, aged 7, speaker ID 1056

Vocal range documented = F₃–A₅, 175–880 Hz

 [Overview of sound sample \(474 sounds\)](#)


Male, aged 8, speaker ID 1057

Vocal range documented = H₃–G₅, 247–784 Hz

 [Overview of sound sample \(394 sounds\)](#)

Male, aged 8, speaker ID 1058

Vocal range documented = A₃–G₅, 220–784 Hz


 [Overview of sound sample \(434 sounds\)](#)

4.1.2 Adult speakers

Women


Female, aged 35, speaker ID 1027

Vocal range documented = C₃–A₅, 131–880 Hz

 [Overview of sound sample \(530 sounds\)](#)


Female, aged 34, speaker ID 1036

Vocal range documented = C₃–A₅, 131–880 Hz

 [Overview of sound sample \(578 sounds\)](#)


Female, aged 25, speaker ID 1039

Vocal range documented = G₃–C₆, 196–1047 Hz









 [Overview of sound sample \(530 sounds\)](#)

Female, aged 28, speaker ID 1088

Vocal range documented = G₃–A₅, 196–880 Hz

 [Overview of sound sample \(482 sounds\)](#)

Men

Male, aged 23, speaker ID 1044Vocal range documented = F₂–G₅, 87–784 Hz [Overview of sound sample \(690 sounds\)](#)Male, aged 31, speaker ID 1045Vocal range documented = A₂–E₅, 110–659 Hz [Overview of sound sample \(594 sounds\)](#)Male, aged 40, speaker ID 1051Vocal range documented = G₂–E₅, 98–659 Hz [Overview of sound sample \(594 sounds\)](#)Male, aged 24, speaker ID 1063Vocal range documented = G₂–G₅, 98–784 Hz [Overview of sound sample \(682 sounds\)](#)**4.1.3 All nonprofessional speakers of the main body of part 1**Children, aged 7 to 10, speaker ID's = 1009, 1034, 1037, 1038, 1054, 1056, 1057, 1058Vocal range documented = F₃–A₅, 175–880 Hz [Overview of sound sample \(3707 sounds\)](#)Women, aged 25 to 35, speaker ID's = 1027, 1036, 1039, 1088Vocal range documented = C₃–C₆, 131–1047 Hz [Overview of sound sample \(2120 sounds\)](#)Men, aged 23 to 40, speaker ID's = 1044, 1045, 1051, 1063Vocal range documented = F₂–G₅, 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, 1088Vocal range documented = F₂–C₆, 87–1047 Hz [Overview of sound sample \(8387 sounds\)](#)

4.2 Professional speakers of the main body of part 1

4.2.1 ST speakers


Women

Female, aged 44, speaker ID = 1046


Professional ranking = 2 (2-BAA)

Total number of sounds = 1059


Vocal range documented for nonstyle productions = E₃–C₆, 165–1047 Hz

 [Overview of sound sample for nonstyle productions \(561 sounds\)](#)

Vocal range documented for style productions = E₃–C₆, 165–1047 Hz

 [Overview of sound sample for style productions \(498 sounds\)](#)

Vocal range documented for all productions = E₃–C₆, 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 nonstyle productions = C₃–G₅, 131–784 Hz

 [Overview of sound sample for nonstyle productions \(521 sounds\)](#)

Vocal range documented for style productions = D₃–G₅, 147–784 Hz

[Overview of sound sample for style productions \(402 sounds\)](#)

Vocal range documented for all productions = C₃–G₅, 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 nonstyle productions = H₂–C₆, 123–1047 Hz

 [Overview of sound sample for nonstyle productions \(665 sounds\)](#)

Vocal range documented for style productions = C₃–A₅, 131–880 Hz

 [Overview of sound sample for style productions \(498 sounds\)](#)

Vocal range documented for all productions = H₂–C₆, 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 nonstyle productions = E₃–G₅, 165–784 Hz

 [Overview of sound sample for nonstyle productions \(481 sounds\)](#)

Vocal range documented for style productions = E₃–G₅, 165–784 Hz

 [Overview of sound sample for style productions \(418 sounds\)](#)

Vocal range documented for all productions = E₃–G₅, 165–784 Hz

 [Overview of entire sound sample \(899 sounds\)](#)


Men

Male, aged 39, speaker ID = 1003


Professional ranking = 2 (2-BAB) and T-AA

Total number of sounds = 1083


Vocal range documented for nonstyle productions = E₂–D₅, 82–587 Hz

 [Overview of sound sample for nonstyle productions \(585 sounds\)](#)

Vocal range documented for style productions = G₂–C₅, 98–523 Hz

 [Overview of sound sample for style productions \(498 sounds\)](#)

Vocal range documented for all productions = E₂–D₅, 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 nonstyle productions = E₂–E₅, 82–659 Hz

 [Overview of sound sample for nonstyle productions \(601 sounds\)](#)

Vocal range documented for style productions = E₂–F₅, 82–698 Hz

 [Overview of sound sample for style productions \(570 sounds\)](#)

Vocal range documented for all productions = E₂–F₅, 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 nonstyle productions = D₂–E₅, 73–659 Hz

 [Overview of sound sample for nonstyle productions \(681 sounds\)](#)

Vocal range documented for style productions = F₂–E₅, 87–659 Hz

 [Overview of sound sample for style productions \(546 sounds\)](#)

Vocal range documented for all productions = D₂–E₅, 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 nonstyle productions = C₂–E₅, 65–659 Hz

 [Overview of sound sample for nonstyle productions \(713 sounds\)](#)

Vocal range documented for style productions = E₂–E₅, 82–659 Hz

 [Overview of sound sample for style productions \(594 sounds\)](#)


Vocal range documented for all productions = C₂–E₅, 65–659 Hz

 [Overview of entire sound sample \(1307 sounds\)](#)

All ST speakers


Women, aged 32 to 51, speaker ID's = 1046, 1048, 1052, 1053

Vocal range documented for all productions = H₂–C₆, 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 = C₂–F₅, 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 = C₂–C₆, 65–1047 Hz

 [Overview of entire sound sample \(8832 sounds\)](#)

4.2.2 CS speakers


Women

Female, aged 26, speaker ID = 1001


Professional ranking = 2 (2-BBAA)

Total number of sounds = 1003

Vocal range documented for nonstyle productions = D₃–A₅, 147–880 Hz

 [Overview of sound sample for nonstyle productions \(553 sounds\)](#)

Vocal range documented for style productions = G₃–C₆, 196–1047 Hz

 [Overview of sound sample for style productions \(450 sounds\)](#)

Vocal range documented for all productions = D₃–C₆, 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 = A₂–A₅, 110–880 Hz

 [Overview of sound sample for nonstyle productions \(553 sounds\)](#)

Vocal range documented for style productions = F₃–F₅, 175–698 Hz

 [Overview of sound sample for style productions \(362 sounds\)](#)

Vocal range documented for all productions = A₂–A₅, 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 nonstyle productions = D₃–C₆, 147–1047 Hz

 [Overview of sound sample for nonstyle productions \(649 sounds\)](#)

Vocal range documented for style productions = F₃–A₅, 175–880 Hz

 [Overview of sound sample for style productions \(460 sounds\)](#)

Vocal range documented for all productions = D₃–C₆, 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 nonstyle productions = E₃–C₆, 165–1047 Hz

 [Overview of sound sample for nonstyle productions \(545 sounds\)](#)

Vocal range documented for style productions = G₃–C₆, 196–1047 Hz

 [Overview of sound sample for style productions \(490 sounds\)](#)

Vocal range documented for all productions = E₃–C₆, 165–1047 Hz

 [Overview of entire sound sample \(1035 sounds\)](#)

Note: Some sounds with reverberation.


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 = D₂–F₅, 73–698 Hz

 [Overview of sound sample for nonstyle productions \(649 sounds\)](#)

Vocal range documented for style productions = G₂–C₅, 98–523 Hz

 [Overview of sound sample for style productions \(501 sounds\)](#)

Vocal range documented for all productions = D₂–F₅, 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 nonstyle productions = G₂–D₅, 98–587 Hz

 [Overview of sound sample for nonstyle productions \(593 sounds\)](#)

Vocal range documented for style productions = A₂–D₅, 110–587 Hz

 [Overview of sound sample for style productions \(460 sounds\)](#)

Vocal range documented for all productions = G₂–D₅, 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 nonstyle productions = A₂–G₅, 110–784 Hz

 [Overview of sound sample for nonstyle productions \(633 sounds\)](#)

Vocal range documented for style productions = A₂–E₅, 110–659 Hz

 [Overview of sound sample for style productions \(507 sounds\)](#)

Vocal range documented for all productions = A₂–G₅, 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 nonstyle productions = D₂–F₅, 73–698 Hz

 [Overview of sound sample for nonstyle productions \(713 sounds\)](#)

Vocal range documented for style productions = G₂–E₅, 98–659 Hz

 [Overview of sound sample for style productions \(546 sounds\)](#)


Vocal range documented for all productions = D₂–F₅, 73–698 Hz

 [Overview of entire sound sample \(1259 sounds\)](#)

All CS speakers


Women, aged 26 to 50, speaker ID's = 1001, 1006, 1023, 1031

Vocal range documented = A₂–C₆, 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 = D₂–G₅, 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 = D₂–C₆, 73–1047 Hz

 [Overview of entire sound sample \(8664 sounds\)](#)

4.2.3 EC speakers


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 = D₃–C₆, 147–1047 Hz

 [Overview of sound sample for nonstyle productions \(625 sounds\)](#)

Vocal range documented for style productions = A₃–A₅, 220–880 Hz

 [Overview of sound sample for style productions \(403 sounds\)](#)

Vocal range documented for all productions = D₃–C₆, 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 = H₂–A₅, 123–880 Hz

 [Overview of sound sample for nonstyle productions \(545 sounds\)](#)

Vocal range documented for style productions = A₃–A₅, 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 = C₃–C₆, 131–1047 Hz

 [Overview of sound sample for nonstyle productions \(633 sounds\)](#)

Vocal range documented for style productions = C₄–C₆, 262–1047 Hz

 [Overview of sound sample for style productions \(428 sounds\)](#)

Vocal range documented for all productions = C₃–C₆, 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 = C₃–C₆, 131–1047 Hz

 [Overview of sound sample for nonstyle productions \(641 sounds\)](#)

Vocal range documented for style productions = C₄–C₆, 262–1047 Hz

 [Overview of sound sample for style productions \(428 sounds\)](#)

Vocal range documented for all productions = C₃–C₆, 131–1047 Hz

 [Overview of entire sound sample \(1069 sounds\)](#)


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 = D₂–H₄, 73–494 Hz

 [Overview of sound sample for nonstyle productions \(577 sounds\)](#)

Vocal range documented for style productions = G₂–G₄, 98–392 Hz

 [Overview of sound sample for style productions \(395 sounds\)](#)

Vocal range documented for all productions = D₂–H₄, 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 = H₂–C₅, 123–523 Hz

 [Overview of sound sample for nonstyle productions \(465 sounds\)](#)

Vocal range documented for style productions = C₃–C₅, 131–523 Hz

 [Overview of sound sample for style productions \(411 sounds\)](#)

Vocal range documented for all productions = H₂–C₅, 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 = F₂–C₅, 87–523 Hz

 [Overview of sound sample for nonstyle productions \(537 sounds\)](#)

Vocal range documented for style productions = G₂–A₄, 98–440 Hz

 [Overview of sound sample for style productions \(411 sounds\)](#)

Vocal range documented for all productions = F₂–C₅, 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 = G₂–C₅, 98–523 Hz

 [Overview of sound sample for nonstyle productions \(505 sounds\)](#)

Vocal range documented for style productions = C₃–C₅, 131–523 Hz

 [Overview of sound sample for style productions \(411 sounds\)](#)


Vocal range documented for all productions = A₂–C₅, 110–523 Hz

 [Overview of entire sound sample \(916 sounds\)](#)

All EC speakers


Women, aged 30 to 56, speaker ID's = 1004, 1005, 1032, 1102

Vocal range documented = H₂–C₆, 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 = D₂–C₅, 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 = D₂–C₆, 73–1047 Hz


 [Overview of entire sound sample \(7828 sounds\)](#)

4.3 Nonprofessional speakers of the side body of part 1 (reference group)

4.3.1 Children


Female, aged 8, speaker ID 1095

Vocal range documented = 15 semitones, A₃–C₅, 220–523 Hz

 [Overview of sound sample \(33 sounds\)](#)


Female, aged 9, speaker ID 1096

Vocal range documented = 15 semitones, A₃–C₅, 220–523 Hz

 [Overview of sound sample \(33 sounds\)](#)


Female, aged 8, speaker ID 1097

Vocal range documented = 15 semitones, A₃–C₅, 220–523 Hz

 [Overview of sound sample \(33 sounds\)](#)


Female, aged 8, speaker ID 1098

Vocal range documented = 15 semitones, A₃–C₅, 220–523 Hz

 [Overview of sound sample \(33 sounds\)](#)


Female, aged 7, speaker ID 1101

Vocal range documented = 15 semitones, A₃–C₅, 220–523 Hz

 [Overview of sound sample \(33 sounds\)](#)


Male, aged 9, speaker ID 1089

Vocal range documented = 15 semitones, A₃–C₅, 220–523 Hz

 [Overview of sound sample \(33 sounds\)](#)


Male, aged 8, speaker ID 1090

Vocal range documented = 15 semitones, A₃–C₅, 220–523 Hz

 [Overview of sound sample \(33 sounds\)](#)


Male, aged 8, speaker ID 1091

Vocal range documented = 15 semitones, A₃–C₅, 220–523 Hz

 [Overview of sound sample \(33 sounds\)](#)


Male, aged 8, speaker ID 1092

Vocal range documented = 15 semitones, A₃–C₅, 220–523 Hz

 [Overview of sound sample \(33 sounds\)](#)

Male, aged 7, speaker ID 1093


Vocal range documented = 15 semitones, A₃–C₅, 220–523 Hz

 [Overview of sound sample \(33 sounds\)](#)

All speakers


Female children, aged 7 to 9, speaker ID's = 1095, 1096, 1097, 1098, 1101

Vocal range documented = 15 semitones, A₃–C₅, 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, A₃–C₅, 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, A₃–C₅, 220–523 Hz


 [Overview of sound sample \(330 sounds\)](#)

4.3.2 Adults

Women


[Female, aged 26, speaker ID 1012](#)

Vocal range documented = 12 semitones, A₃–A₄, 220–440 Hz

 [Overview of sound sample \(25 sounds\)](#)


[Female, aged 24, speaker ID 1016](#)

Vocal range documented = 12 semitones, A₃–A₄, 220–440 Hz

 [Overview of sound sample \(25 sounds\)](#)


[Female, aged 20, speaker ID 1020](#)

Vocal range documented = 12 semitones, A₃–A₄, 220–440 Hz

 [Overview of sound sample \(25 sounds\)](#)


[Female, aged 26, speaker ID 1021](#)

Vocal range documented = 12 semitones, A₃–A₄, 220–440 Hz

 [Overview of sound sample \(25 sounds\)](#)


[Female, aged 28, speaker ID 1040](#)

Vocal range documented = 12 semitones, A₃–A₄, 220–440 Hz

 [Overview of sound sample \(25 sounds\)](#)


[Female, aged 27, speaker ID 1041](#)

Vocal range documented = 12 semitones, A₃–A₄, 220–440 Hz

 [Overview of sound sample \(25 sounds\)](#)


[Female, aged 29, speaker ID 1066](#)

Vocal range documented = 12 semitones, A₃–A₄, 220–440 Hz

 [Overview of sound sample \(25 sounds\)](#)


[Female, aged 37, speaker ID 1071](#)

Vocal range documented = 12 semitones, A₃–A₄, 220–440 Hz

 [Overview of sound sample \(25 sounds\)](#)


[Female, aged 34, speaker ID 1080](#)

Vocal range documented = 12 semitones, A₃–A₄, 220–440 Hz

 [Overview of sound sample \(25 sounds\)](#)

[Female, aged 25, speaker ID 1081](#)


Vocal range documented = 12 semitones, A₃–A₄, 220–440 Hz

 [Overview of sound sample \(25 sounds\)](#)

Men


Male, aged 47, speaker ID 1011

Vocal range documented = 12 semitones, C₃–C₄, 131–262 Hz

 [Overview of sound sample \(25 sounds\)](#)


Male, aged 21, speaker ID 1013

Vocal range documented = 12 semitones, C₃–C₄, 131–262 Hz

 [Overview of sound sample \(25 sounds\)](#)


Male, aged 24, speaker ID 1015

Vocal range documented = 12 semitones, C₃–C₄, 131–262 Hz

 [Overview of sound sample \(25 sounds\)](#)


Male, aged 27, speaker ID 1022

Vocal range documented = 12 semitones, C₃–C₄, 131–262 Hz

 [Overview of sound sample \(25 sounds\)](#)


Male, aged 33, speaker ID 1024

Vocal range documented = 12 semitones, C₃–C₄, 131–262 Hz

 [Overview of sound sample \(25 sounds\)](#)


Male, aged 18, speaker ID 1073

Vocal range documented = 12 semitones, C₃–C₄, 131–262 Hz

 [Overview of sound sample \(25 sounds\)](#)


Male, aged 52, speaker ID 1075

Vocal range documented = 12 semitones, C₃–C₄, 131–262 Hz

 [Overview of sound sample \(25 sounds\)](#)


Male, aged 35, speaker ID 1078

Vocal range documented = 12 semitones, C₃–C₄, 131–262 Hz

 [Overview of sound sample \(25 sounds\)](#)


Male, aged 25, speaker ID 1084

Vocal range documented = 12 semitones, C₃–C₄, 131–262 Hz

 [Overview of sound sample \(25 sounds\)](#)

Male, aged 25, speaker ID 1085


Vocal range documented = 12 semitones, C₃–C₄, 131–262 Hz

 [Overview of sound sample \(25 sounds\)](#)

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, A₃–A₄, 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, A₃–A₄, 131–262 Hz

 [Overview of sound sample \(250 sounds\)](#)

All nonprofessional adults, aged 18 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

Vocal range documented = 12 semitones, A₃–A₄, 131–440 Hz


 [Overview of sound sample \(500 sounds\)](#)

4.4 Summary

4.4.1 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 = F₂–C₆, 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 = C₂–C₆, 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 = D₂–C₆, 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 = D₂–C₆, 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 = C₂–C₆, 65–1047 Hz

 [Overview of entire sound sample \(33711 sounds\)](#)

4.4.2 Side body

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, A₃–C₅, 131–523 Hz

 [Overview of sound sample \(830 sounds\)](#)

5 Parts 2 to 5 of the Zurich Corpus – Details of method and links to sound samples

With regard to the parts 2 to 5 of the corpus, some information is given below in addition to the description in Chapter 1.3, including sound links.

Concerning the recordings of speech extracts documented in part 2 of the published corpus, information about speakers, speech and speech context as well as references are given in the speaker information and the sound comment fields in the corpus. (For further details, see Maurer, 2024, Chapters 2.4 and M2.4.) In the corpus, the entire sound sample of the speech extracts is assigned to the category A-BB:

<https://zhcorpus.org/v2/db/authors?cat=A-BB&info=1>

Concerning the recordings of syllables and minimal pairs documented in part 3 of the published corpus – at the date of its first publication –, two subsamples of minimal pairs produced by two women are provided. (For further details, see Maurer, 2024, Chapters 2.3 and M2.3.) Additional samples will be added and described in the Additions section of the corpus. In the corpus, the entire sound sample of the speech extracts is assigned to the category A-BC:

<https://zhcorpus.org/v2/db/authors?cat=A-BC&info=1>

Concerning manipulated natural, resynthesized/synthesised and filtered sounds documented in part 3 of the published corpus, as mentioned, they relate to experiments described in Maurer (2024). Additional sounds and sound types will be added and described in the Additions section of the corpus. In the corpus, the entire sample of these sounds is assigned to the category A-BD:

<https://zhcorpus.org/v2/db/productionMatrix?cat=A-BD&info=1>

As also mentioned, part 5 of the work database comprises additional natural sounds that either do not belong to the sound selection of part 1 (because of their ranking status) or do not belong to the sounds of parts 2 to 4. In the published corpus – at the date of its first publication –, these sounds again relate to documentations and experiments of Maurer (2024). Additional sounds will be added and described in the Additions section of the corpus. In the corpus, the entire sample of these sounds is assigned to the category A-BE:

<https://zhcorpus.org/v2/db/productionMatrix?cat=A-BE&info=1>

6 Software tools implemented into the corpus

Four software tools were developed by Christian d'Heureuse (2018, 2019a, 2019b, 2022) and are implemented as web-based applications into the corpus: A reviewed and adapted version of the Klatt synthesiser (KlattSyn tool), a sinusoidal synthesiser (SinSyn tool), a harmonic analyser and (re-)synthesiser (HarmSyn tool) and a sound filter tool (SpecFilt tool). Short descriptions of the tools are given in Maurer (2024, pp. 47-50). For further details, source codes, demo versions and related comments, see the below links in the References section.

For the functionality integrated into the corpus and the corresponding graphical user interface, refer to the help 'h' in the user interface (line of the tools below a sound spectrum; see right). For default parameters, see the parameter forms of the tools. Note also that some parameter fields have a tooltip.

References:

- d'Heureuse, C. (2018). SinSyn – Sinusoidal synthesizer [browser-based web application]. GitHub repository. <https://github.com/chdh/sin-syn>. Retrieved 20 Nov. 2022.
- d'Heureuse, C. (2019a). KlattSyn – Klatt formant synthesizer [browser-based web application]. GitHub repository. <https://github.com/chdh/klatt-syn>. Retrieved 20 Nov. 2022.
- d'Heureuse, C. (2019b). HarmSyn – Harmonic analyzer and synthesizer [browser-based web application]. GitHub repository. <https://github.com/chdh/harm-syn>. Retrieved 20 Nov. 2022.
- d'Heureuse, C. (2022). SpecFilt – Spectral filter tool [browser-based web application]. GitHub repository. <https://github.com/chdh/spec-filt>. Retrieved 20 Nov. 2022.

7 Accessibility and terms of use

The database can be downloaded. The download consists of the database information in TXT format and the recordings in WAV format. However, restrictions apply. For details and a download request, please refer to the link "Terms of use" in title page of this website.

Appendix

A1 Fundamental frequencies investigated and notation of musical 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) and further adapted for the professional speakers investigated.

Reference: Bunch, M., & Chapman, J. (2000). Taxonomy of singers used as subjects in scientific research. *Journal of Voice*, 14(3), 363–369. Carpenter, A., & Morton.

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

./ (continuation see next page)

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 tests

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. Indications in the screen are as follows.

Listener: LDM 1932-A Votes:1 Open:24 Total:25 211588

Back Mark Play Delay 0 Index

u o o1 a-a1 e i ä ö ü

u-o o-o1 o1-a Schwa e-i ä-ö ö-ü

steigend hoch-hoch a1-ä e-ä i-ü ä-e ö-e ü-i

flach

fallend tief-tief a1 e-ö ä-a1 ü-e

no / 1 yes / 2 a e-ü other (n.s.) Text

Test information

Header line: Listener ID (e.g. LDM), identification test number (e.g. 1032-A), number of sounds already identified ("Votes"), number of sounds to identify ("Open"), total number of sounds of the subseries ("Total"), actual sound ID number of the sound in the corpus.

Vowel quality recognition tests

First row of buttons:

- "Back"; using this button, the listener could go back to the previous sound (only) to allow correction of typing errors; however, it was not possible to go back further in a test.
- "Mark"; this button allowed the listener to mark a sound to indicate that the investigators should check the sound characteristics again (e.g. for background noise).
- "Play"; with this button, the listener could repeat the playback of the sound; there was no restriction for the repeat playback.

"Delay"; when the listener clicked on a vowel button to assign a perceived vowel quality, the next sound was automatically played; after the click, however, a delay time (in seconds) was added, which the listener could set via the 'Delay' field.

"Index"; this button allowed the listener to exit the subtest and return to the subtest list.

Buttons in square format, button 'other (n.s.) and field "Text": With the buttons in square format, the listener could assign a single perceived vowel quality or any combination of two

neighbouring qualities. With the 'other (n.s.)' button, the listener could indicate that no vowel quality was recognised (result = 'not specified' or 'ns'). With the 'Text' field, the listener could indicate that they recognised a vowel quality or vowel qualities other than those specified in the square buttons (e.g. an uncertain recognition of /u/ or /i/).

Pitch level recognition tests

Three additional buttons for pitch recognition tests (lower left quadrant) were used for specific experiments, in which a first reference sound and a second test sound as one test item were presented:

- "Steigend" ("rising"); using this button, the listener recognised the second sound at a higher pitch level than the first sound.
- "Fallend" ("falling"); using this button, the listener recognised the second sound at a lower pitch level than the first sound.
- "hoch–hoch / flach / tief tief" ("high high / flat / low low"); using this button, the listener recognised the second and first sound at a similar pitch level.

Double-vowel and double-pitch recognition tests

Two additional buttons for double-vowel or double-pitch recognition tests (lower left quadrant) were used for specific experiments in which single sounds as test items were presented:

- "no / 1"; using this button, the listener recognised only one vowel quality or only one pitch level.
- "yes / 2"; using this button, the listener recognised two vowel qualities or two pitch levels.