The Zurich Corpus of Vowel and Voice Quality

Version 2
zhcorpus.org

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

Handbook

(Last update 2024-07-31)

Contents

Introductory note ... 3

1. The Zurich Corpus ... 4
1.1 Background ... 4
1.2 The work version of the database ... 4
1.3. Published Version 1 of the Corpus ... 7
1.4 Published version 2 of the corpus ... 7

2 Terms and abbreviations ... 9
2.1 Entire listing ... 9
2.2 Speakers and styles ... 10
2.3 Vowel qualities and vowel notation ... 10
2.4 Production parameters ... 10
2.5 Sound status and sound ranking ... 11

3 Part 1 of the Zurich Corpus – Details of method ... 13
3.1 Speakers and utterances ... 13
3.2 Recordings ... 17
3.3 Acoustic analysis ... 20
3.4 Listening test ... 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 Non-professional speakers of the main body of part 1 ... 25
4.1.1 Children ... 25
4.1.2 Adult speakers ... 25
4.1.3 All non-professional 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 Non-professional 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 and links to sound samples ... 37

6 Software tools implemented into the corpus ... 38

7 Accessibility and terms of use ... 41

Appendix ... 41
A1 Fundamental frequencies investigated and notation of musical C major scale ... 41
A2 Taxonomy of professional actresses/actors and singers ... 42
A3 Screen of the listening test ... 43

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

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). Online on https://zhcorpus.org/v2/doc/MaurerEtAl2018.pdf

Maurer, D. (2016): *Acoustics of the Vowel – Preliminaries*. Peter Lang.

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_o, 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_o 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_o, 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 dor 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_o 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_o 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_o 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_o ranges and at highlighting the significance of extensive f_o 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_o 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.

Below, details on all five parts of the sound database (work version, unpublished, including all recordings made until 2022) are given, followed by a description of the first two versions published online.

Part 1 – Natural vowel sounds, produced with a systematic variation of basic production parameters

Part 1 of the corpus has a double structure: The main body consists of sounds of a large-scale investigation and documentation of the long Standard German vowels /i, y, e, ø, ɛ, a, o, u/, the sounds produced with extensive variation of basic production parameters by 16 nonprofessional and 24 trained and professionally active speakers and singers (hereafter nonprofessionals and professionals). For all speakers, a read text, and for professional speakers and singers, one or several songs are also included. The side body consists of reference recordings of the same set of vowel sounds produced by 30 nonprofessionals, with no production parameter variations except f_o variation within an everyday speaking range. A read text is also included.

Details of speakers, utterances, recordings, acoustic analysis, listening test and sound selection for publication are given in Chapter 3.

Part 2: Extracts of speech and singing

Part 2 of the corpus consists of speech extracts produced by speakers without formal vocal training, by politicians, journalists and TV hosts, and by professionally trained speakers from the field of the performing arts (actresses/actors and singers). The utterances were either recorded in person (live recordings conducted by the author, with consent for publication given by the speakers) or extracted from taped TV shows, Internet content or DVDs/CDs. The extracts aim to document and highlight the observable f_o range found for everyday speech and for speech and singing in the field of the performing arts.

Part 3: Syllables and minimal pairs

Part 3 of the corpus consists of syllables and minimal pairs produced by single speakers at various f_0 levels. It includes sounds collected in the context of studies on the intelligibility of vowel sounds produced at middle and high f_0 levels (see Maurer et al., 2014; Friedrichs et al., 2015a, 2015b, 2017) and sounds of selected professional speakers recorded in the general context of the building up of the corpus.

References:

Friedrichs, D., Maurer, D., & Dellwo, V. (2015a). The phonological function of vowels is maintained at fundamental frequencies up to 880 Hz. *The Journal of the Acoustical Society of America*, *138*(1), EL36–EL42.

Friedrichs, D., Maurer, D., Suter, H., & Dellwo, V. (2015b, August): Vowel identification at high fundamental frequencies in minimal pairs. In *ICPhS* (0434, 1–4).

Friedrichs, D., Maurer, D., Rosen, S., & Dellwo, V. (2017). Vowel recognition at fundamental frequencies up to 1 kHz reveals point vowels as acoustic landmarks. *The Journal of the Acoustical Society of America*, *142*(2), 1025–1033.

Maurer, D., Mok, P., Friedrichs, D., & Dellwo, V. (2014). Intelligibility of high-pitched vowel sounds in the singing and speaking of a female Cantonese Opera singer. In Li, H., Meng, H. M., Ma, B., Chng, E. S., & Xie, L. (Eds.)**,** *15th Annual Conference of the International Speech Communication Association, INTERSPEECH 2014* (pp. 2132–2133). Curran Associates, Inc. (Materials: https://is2014.phones-and-phonemes.org/en/. Retrieved 20. Nov. 2022.)

Part 4: Manipulated natural sounds, resynthesised and synthesised sounds

Part 4 of the database consists of manipulated natural sounds and of resynthesised and synthesised sounds investigated in the context of specific experiments, that is, LP- and HP-filtered sounds as well as resynthesised and synthesised sounds using either a Klatt synthesiser or a sinusoidal synthesiser or a harmonic synthesiser (for the corresponding tools, see below).

Part 5: Miscellaneous

Part 5 of the corpus is comprised of additional natural sounds that do not belong to the sound sample of the previous parts. They consist of the following categories: (1) Sounds produced by speakers who were not able to satisfactorily produce vowel sounds of sufficient quality for all investigated production parameters during the recording sessions; (2) various sounds of the vowel /ɔ/ that, initially, were intended to be part of the sample of Part 1 but proved to be too difficult to produce for some speakers (above all for nonprofessionals; therefore, in the course of creating the Zurich Corpus, we decided not to pursue further recordings of this vowel while still keeping the sounds we had already recorded as part of the miscellaneous sounds of this fifth part); (3) sounds produced by some of the speakers in sVsV context at f_0 levels in a lower frequency range not corresponding to the standard range of the sounds of Part 1; (4) duplicates (entire sounds or sound nuclei) of natural sounds used for specific experiments and related vowel and/or pitch recognition tests. In addition, for some speakers, a few glissandi and schwa sounds were also recorded.

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.

Acoustic analysis and listening test

Details on the acoustic analysis and the standard listening test conducted are given below.

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)
CS	contemporary singing style (generic term, including the substyles contemporary musical theatre, pop, rock, and jazz)
D	standard recording, but deficient (details see individual sound comments)
EC	European classical singing style
F1, F2, ...	formant frequencies
f_0	fundamental frequency
hgh	high vocal effort
hp	HP filtered.
K	recordings made in a concert hall
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	produced in minimal pair 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
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 (s = consonant /s/, V = /vowel/)
T	test sounds (recordings made/sounds produced for specific investigations)
tr	text read
ts	text sung
v	voiced phonation
V	vowel sound produced in isolation
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 non-professional 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 /ɑ/, 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 (synthesis, different synthesis methods)
hp	HP filtered
lp	LP filtered
nc	extracted sound nucleus

Vowel context and related abbreviations:

V	in isolation
sVsV	in s-V-s-V context (s = consonant /s/, V = /vowel/)
mp	in minimal pair context
text	in read 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.

3.1 Speakers and utterances

Overview

Tables 2 and 3 show the speakers and production parameters of systematic investigation as documented in the first part of the corpus (extract of Maurer, 2024).

Table 2. The Zurich Corpus, Version 2, Part 1, systematic compilation of one sound per production parameter configuration: Speakers. [C01-01-T02]

Speaker group	Children		Adults		Speakers
	f	m	f	m	(total)
Main body					
Non-professionals	4	4	4	4	16
ST actresses/actors	–	–	4	4	8
CS singers	–	–	4	4	8
EC singers	–	–	4	4	8
Side body					
Non-professionals	5	5	10	10	30
Total	9	9	26	26	70

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

Table 3. The Zurich Corpus, Version 2, Part 1, systematic compilation of one sound per production parameter configuration: 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

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). Colum 4 = intended f_o level variation (musical scale = according to C-major scale; upper scale and reference f_o, see text). Column 5 = production style (N = nonstyle, S = style of ST, CS or EC).

Speakers of extensive investigation (main body)

As shown in Table 2, 16 non-professionals (8 children, aged 7 to 10, and 8 adults, aged 23 to 40, gender balanced) and 24 professionals (adults, aged 25 to 56, gender balanced) with no report of hearing impairment were investigated concerning utterances with extensive varying basic production parameters. The professional group is comprised of 8 ST actresses/actors, 8 CS singers (including the substyles contemporary musical theatre, pop, and jazz), and 8 EC singers (2 sopranos, 2 mezzo-sopranos, 2 tenors, and 2 baritones).

Non-professionals were selected according to two criteria: a minimal vocal range for vowel production of 2 octaves (24 semitones) for adults and 1.5 octaves (19 semitones) for children, with vowels recognisable over a range of 15 semitones in minimum for both adults and children. The children's age was set at 7 to 10 years for the following reasons: The maximal age of 10 was set to control for voice changes that happen in puberty, the minimal age of 7 was set to ensure that the children are mature enough to follow instructions.

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 (2000; see Appendix), 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.

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

Utterances of extensive investigation (main body)

As shown in Table 3, the speakers of extensive investigation produced sustained sounds of the eight long Standard German vowels /i–y–e–ø–ɛ–a–o–u/ with varying basic production parameters for phonation (voiced, breathy, creaky, whisper), vocal effort (medium, low, high, shouted), vowel context (V and sVsV), and f_0 (monotonous pitch levels according to C-major scale, full f_0 range according to assigned register/voice range). All utterances were made by the speakers as non-professional (nonstyle) 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 nonstyle utterances, the professionals were also asked to produce the same set of voiced sounds in their own respective singing/performance style and for an f_0 range that reflects their artistic style, with a corresponding variation of vocal effort, vowel context and f_0. Thus, vowel sound 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 produced with medium vocal effort on a higher f_0 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 f_o range, very high vocal effort and whispering.

For sound duration, see below.

The non-professionals also read a reference text on the spot ("Nordwind und Sonne", see the Handbook of the International Phonetic Association, 1999, pp. 88–89) and were asked to sing a song in German. For two children, additional songs were also recorded. The professionals read the same text on the spot in nonstyle as well as in style mode and were asked to sing a prepared song in German in their respective singing style. For some professionals, additional songs in Italian, English or French were also recorded.

Reference speakers (side body)

Vowel production with very limited f_o variation of 30 native Swiss German non-professional speakers (10 children, aged 7–9, and 20 adults, aged 18–52, gender balanced) were also investigated (see Table 2). The speakers were selected according to their native Swiss German dialect (only speakers of Swiss German dialects from the eastern Swiss midland, mainly the Canton Zurich region, were included), their command of speaking and pronouncing Standard German (primary language used in schools in Switzerland) and their ability to produce recognisable 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 speakers produced sustained sounds of the eight long Standard German vowels 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, firstly, investigate sounds that mirror a common range of f_o contours of everyday speech that has to be considered when investigating vowel sounds, that is, a range which can be observed with no imperative and thus pronounced register change for 262–523 Hz for children, 220–440 Hz for women and 131–262 Hz for men, respectively, and secondly, to allow for a comparison of sounds on different and similar f_o for different age- and gender-related speaker groups. In addition, the speakers were asked to spontaneously read out loud the reference text (see above).

This sample of reference speakers and utterances – limited to native speakers of Swiss German dialects spoken in a restricted geographical region – was collected to show that, in comparison to these speakers, all speakers of extensive investigation (with less regional restriction of speaker origin and, in part, with professionally trained voices) generally show comparable vowel pronunciation both in terms of acoustic characteristics and vowel recognition, given nonstyle mode, corresponding f_o and vocal effort levels of sound production.

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 German native speakers (or if not, good command of German phonetic articulation on stage; see below for 'professional speakers with native languages other than German'), vocal range of 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 non-professional speaker's group (main body, adults and children), our requirements were as follows: native speakers of German (including Swiss German), vocal range of 2 octaves (adults) and 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: native speakers of Swiss German dialects from the eastern Swiss midland, ability to produce recognisable vowel sounds on a specific pitch over an f_0 range of 15 semitones in minimum, 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 native languages other than German: As mentioned, four professional singers are not native speakers of German 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, many speakers exhibited severe 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 in order 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 non-professional speakers (main body), vowel sounds with creaky phonation were not included in the standard recordings of investigation. However, during the recordings, a few non-professional 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 & Kiær 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.

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_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 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_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 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 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 more details below). 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 non-professional 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 about 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 occurred. Despite clear recording instructions, some non-professional 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, in general, the sound duration of vowel nuclei (excluding on- and offset) 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 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_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 middle 0.3 sec of the first or the second vowel sound, depending on their duration. Concerning the allocation of the 03. 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 duration was ≥ 0.5 sec or, if not, its duration was longer than the duration of the second vowel. If this does not apply, the second vowel nucleus was assigned. Subsequently, the corresponding recording was visually crosschecked. If the automatic estimation failed, the nucleus was manually set by the investigator (first author). – For sVsV context and 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 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_0 contour, the spectrogram and the LTAS. Numerical average values of f_0 and formant patterns were added to the sound information.

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. Graphic representations and numerical values were visually crosschecked for accuracy of cuts, assignment of 0.3 sec vowel nucleus and calculated 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.

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 background noise for sound recording with high gain level and low vocal effort (see above).

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 sounds, the amplitudes of the sounds are normalised.

Sound status

All sounds of the first part of the corpus, which were recorded on the basis of 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).

3.4 Listening test for vowel quality recognition

Listeners

Five phonetic expert listeners (professionally trained singers or actors or voice teachers) performed listening tests for all 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) in order to exclude hearing impairment.

Test procedure

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 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 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.

The vowel /ɔ/ was included in the listening test because the perceptual distance /a–o/ 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/.

Further details

Type and number of listeners: 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, in order 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.

Vowel recognition rate: 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).

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, 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 tool for the selection of the most representable sound 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 ≥ 60%.
  - r1 = the sound with the highest vowel recognition rate ≥ 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 ≥ 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 r2 status.
  - r4 = the sound(s) with the a vowel recognition rate < 60 % is/are assigned r2 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 ≥ 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 recogntion 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 Non-professional speakers of the main body of part 1

4.1.1 Children

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)
Note: Some sounds contain reverberation.

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

Male, aged 8, speaker ID 1054
Vocal range documented = G_3–F_5, 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 = 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)

4.1.2 Adult speakers

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)

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)

4.1.3 All non-professional speakers of the main body of part 1

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 non-professional 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)

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 = E_3–C_6, 165–1047 Hz
Overview of sound sample for nonstyle 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 nonstyle productions = C_3–G_5, 131–784 Hz
Overview of sound sample for nonstyle 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 nonstyle productions = H_2–C_6, 123–1047 Hz
Overview of sound sample for nonstyle 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 nonstyle productions = E_3–G_5, 165–784 Hz
Overview of sound sample for nonstyle 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)

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 = E_2–D_5, 82–587 Hz
Overview of sound sample for nonstyle 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 nonstyle productions = E_2–E_5, 82–659 Hz
Overview of sound sample for nonstyle 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 nonstyle productions = D_2–E_5, 73–659 Hz
Overview of sound sample for nonstyle 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 nonstyle productions = C_2–E_5, 65–659 Hz
Overview of sound sample for nonstyle 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)

All ST 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)

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 = D_3–A_5, 147–880 Hz
Overview of sound sample for nonstyle 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 nonstyle 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 nonstyle productions = D_3–C_6, 147–1047 Hz
Overview of sound sample for nonstyle 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 nonstyle productions = E_3–C_6, 165–1047 Hz
Overview of sound sample for nonstyle 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)
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 = D_2–F_5, 73–698 Hz
Overview of sound sample for nonstyle 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 nonstyle productions = G_2–D_5, 98–587 Hz
Overview of sound sample for nonstyle 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 nonstyle productions = A_2–G_5, 110–784 Hz
Overview of sound sample for nonstyle 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 nonstyle productions = D_2–F_5, 73–698 Hz
Overview of sound sample for nonstyle 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)

All CS 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)

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 = D_3–C_6, 147–1047 Hz
Overview of sound sample for nonstyle 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 nonstyle 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 nonstyle 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 nonstyle 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)

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 nonstyle 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 nonstyle 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 nonstyle 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 nonstyle 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)

All EC 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)

4.3 Non-professional 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, 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)

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)

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)

4.3.2 Adults

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)

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)

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 non-professional 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, A_3–A_4, 131–440 Hz
Overview of sound sample (500 sounds)

4.4 Summary

4.4.1 Main body

All non-professional 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)

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, A_3–C_5, 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, 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, resynthesised and synthesised sounds documented in part 3 of the published corpus, as mentioned, they relate to experiments described in Maurer (2024). (Except are sounds which are or 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 do either not belong to the sound selection of part 1 (because of their ranking status) or not belong to the sounds of parts 2 to 4. In the published corpus, these sounds again relate to documentations and experiments of Maurer (2024). (Except are sounds which are or 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). The features of the tools are outlined in short below. 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 its graphical user interface, refer to the help in the user interface (line of the tools below a sound spectrum). For default parameters, see the parameter forms of the tools. Note also that each parameter field has 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/spect-filt. Retrieved 20 Nov. 2022.

KlattSyn

The KlattSyn tool is a redevelopment of the classic Klatt cascade-parallel formant synthesiser that allows for a source-filter sound synthesis. The concept of the synthesiser is elaborated in Klatt (1980) and Klatt and Klatt (1990), and the review and adaptations made for the newly developed KlattSyn tool is described in d'Heureuse (2019a).

In the Zurich corpus, for each single sound, a link to the KlattSyn tool is listed. If activated, the parameter form for the synthesiser is displayed in which, automatically, the results acoustic analysis for a sound are inserted (average f_0 and average frequencies, levels and bandwidths of the formants; standard analysis of the corpus, with age- and gender-related default parameters for LPC analysis). If needed, these parameters can be edited manually. On the basis of these parameters, the (re-)synthesised sound and its average spectrum and vocal transfer function is displayed as assessed by the Klatt synthesiser, and the sound can be played back and saved.

Thus, for a single vowel sound and its calculated average values for f_0 and formants, the KlattSyn tool allows for a direct resynthesis to assess the perceptual relevance of an estimated *F*-pattern, that is the perceptual correspondence of the original natural sound and the resynthesised replica, the replica being based on the
results of formant analysis. Further, the tool allows for an investigation of the perceptual significance of changes in the synthesis parameter setting, which is of particular importance regarding the question of the effect of source variation with maintained filter pattern.

If the parameter form is reset, the Klatt synthesis can be configured for any setting of parameters.

For a direct Klatt resynthesis (based on calculated f_0 and formants) without display of the parameter form, a corresponding resynthesis play button and a related parameter field for f_0 are given in the figure legend of a sound.

SinSyn

The SinSyn tool (d'Heureuse, 2018) is a tool for sound synthesis based on any series of sine waves (frequencies, amplitudes and phases).

In the Zurich Corpus, for each single sound, a link to the SinSyn tool is listed. If activated, the parameter form of the synthesiser is displayed in which, automatically, the results of LPC analysis for the first three formants (frequencies and levels; standard analysis of the corpus, with age- and gender-related default parameters for LPC analysis) are taken as *S*1–*S*2–*S*3 patterns according to the concept of so-called sinewave vowel sounds as discussed in the literature. Concerning the matter of the present investigation, the SinSyn tool allows for a three-sinewave synthesis that relates to an estimated pattern of the first three formants (their frequencies) of a natural sound in order to assess the perceptual correspondence of vowel quality of the two sounds.

Besides, sinewave synthesis allows for a resynthesis based on a series of sinusoids mirroring the harmonic spectrum of a natural reference sound. Further, sinewave synthesis also allows for an investigation of the vowel sound beyond the framework of source and filter, with any number and with or without a harmonic relation of the sinusoids. However, the quality of the synthesised sounds is very limited.

For a direct sinewave synthesis (related to the calculated first three formants) without display of the parameter form, a corresponding synthesis play button is given in the figure legend of a sound.

HarmSyn

The HarmSyn tool (d'Heureuse, 2019b) is based on an analysis and (re-)synthesis algorithm for quasi-periodic sounds. The sound analysis part allows for the calculation of the dynamic course of f_0 and the harmonic spectrum (frequencies and amplitudes). The sound synthesis part allows for either a direct resynthesis in terms of calculating back a sound on the basis of acoustic analysis of the natural reference sound or a synthesis based on a manipulation of the analysed harmonic spectrum (deletion of single harmonics). In the command line mode, in addition, the synthesis based on a manipulation of the analysed harmonic spectrum includes the option of an alteration of harmonic amplitudes.

If playback functionality is enabled for a sound in the Zurich Corpus, a link to the HarmSyn tool is listed. If activated, the parameter form of the tool is displayed subdivided into the two parts of analysis and of synthesis. For an acoustic analysis, the corresponding parameters can be selected. For a (re-)synthesis, the series of harmonics and their levels are inserted into the form after the analysis was performed. If needed, the indications of the harmonic spectrum can be edited (enabling/disabling individual harmonics or, in the command line version, amplifying/attenuating their level). Subsequently, the (re-)synthesis can be performed, and the resulting sound can be played back and saved.

Thus, for a single vowel sound with approximately periodic sound characteristics and with its calculated dynamic course of f_0 and the harmonic spectrum, the HarmSyn tool allows for the resynthesis of the sound in order to assess the perceptual correspondence of the natural reference sound and its resynthesised replica. Further, the tool allows for an investigation of the perceptual correlate of selected harmonics and/or, in the command line version, of increasing or

decreasing harmonic levels in sound synthesis, which is of particular importance regarding the question of the spectral representation of vowel quality.

Noteworthy, the HarmSyn tool allows for an investigation of the vowel sound beyond the framework of source and filter and it is able to produce a very good sound quality of the (re-)synthesised sounds even for a highly reduced number of harmonics. With regard to sound quality, it by far surpasses the Klatt and sinewave synthesisers.

SpecFilt

The SpecFilt tool (d'Heureuse, 2022) is a tool for LP, BP and HP sound filtering including the option of filtering based on a custom free-form filter curve. It allows for the calculation of a Fourier spectrum of a sound or part of it, for spectral filtering and for subsequent inverse Fourier Transform and for final playback and saving of the filtered sound or sound part.

If playback functionality is enabled for a sound in the Zurich Corpus, for each single sound, a link to the SpecFilt tool is listed. If activated, the parameter form of the tool is displayed in which, automatically, the sound is inserted. Filter parameters can be set, and the correspondingly filtered sound can be played back and downloaded. (Note that special attention should be given to the window function parameter.)

Concerning the matter of the present investigation, the SpecFilt tool allows for the verification of the LP and HP filter experiments and their results presented and for a further exploration of the effect of sound filtering on recognised vowel quality, above all in the context of the foreground–background thesis put forward in this treatise. This kind of sound manipulation is again of particular importance regarding the question of the spectral representation of vowel quality.

For direct sound filtering without display of the parameter form, a corresponding resynthesis play button and related parameter fields for filter types and cutoff frequencies are given in the figure legend of a sound.

7 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

For 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; 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

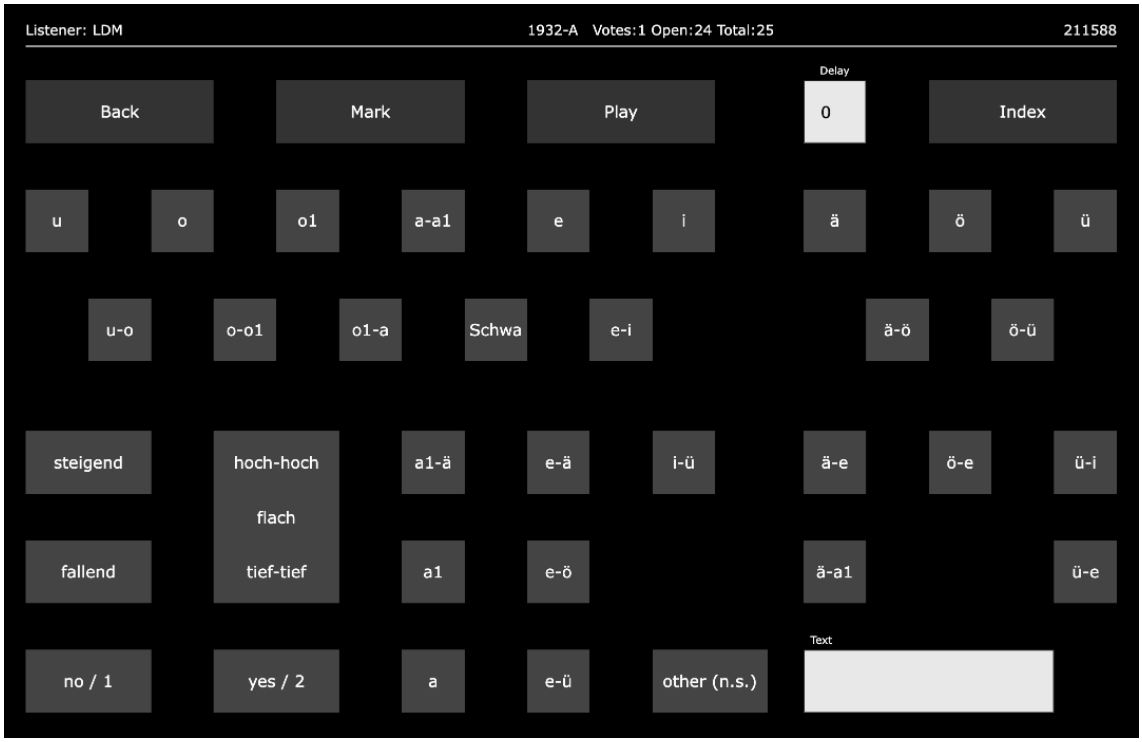
3-C EC singers
- 3-CA Opera
  - 3-CAA Actor/actress in a leading role
  - 3-CAB Actor/actress in a supporting role
- 3-CB Concert/oratorio/recital principal

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 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.



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 line of buttons:

- “Back”; using this 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 a test.
- “Mark; using this button, the listener could mark a sound in order to indicate that the investigators should double-check the sound characteristics (e.g. for background noise).
- “Play”; using this button, the listener could repeat the playback of the sound; no restriction was made for the playback repetition.
- “Delay”; 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 (in sec) was added which could be set by the listener using the “Delay” field.
- “Index”; using this button, the listener could quit the subtest and go back to the subtest listing.

Buttons in square format, button “other (n.s.) and “Text” field: 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 no vowel quality was recognised (result = “not specified” or “ns”). Using the field “Text”, the listener could indicate a recognition of a vowel quality or of vowel qualities other than indicated by 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 on a higher pitch level than the first sound.
- “Fallend” (“falling”); using this button, the listener recognised the second sound on 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 on 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.