The Zurich Corpus of Vowel and Voice Quality

Version 2

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

About

(Extract of the handbook, Chapter 1)

Background

Besides the 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. (For further details, examples and references, see Maurer et al., 2018.)

However, in general, these existing samples or databases either present sounds produced by speakers with medium vocal effort at particular f_0, or they compare sounds of speakers related to only two different production parameters, e.g., voiced and whisper phonation, or V and CVC context, or voiced with varying vocal effort, or voiced with varying f_0 in singing, etc. (for references, see Maurer et al., 2018). To the best of our knowledge, no database exists that includes an extensive and combined variation of basic production parameters such as phonation type, vocal effort, f_0, and vowel context for the sounds of each single documented speaker. Therefore, hitherto, we did not have phenomenological and descriptive references at our disposal that allow for the acquisition of a comprehensive knowledge and 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.

The work 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the Zurich Corpus of Vowel and Voice Quality – in short the Zurich Corpus – was created in the form of an extensive unpublished sound database (hereafter work database or work version of the corpus), with selected, smaller versions thereof published online with open access.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 being published in the context of a treatise entitled Acoustics of the Vowel – Indices (see Maurer 2023).

The entire sound database, that is, the work version of the corpus 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 from stage performances and films
- Part 3 Syllables and minimal pairs produced by single speakers at different f_0 levels
- Part 4 Manipulated natural, resynthesised and synthesised sounds
- Part 5 Miscellaneous

The first and most extensive part addresses the question of observable acoustic characteristics of vowel sounds. The second part addresses the question of the observable ranges of f_0 in intelligible speech and singing. The third part documents vowel sounds produced in the specific context of syllables and minimal pairs by single speakers at various f_0 levels. The fourth part documents sounds investigated in the context of different experiments related to sound filtering, resynthesis and synthesis. The fifth part consists of miscellaneous sounds that were cast aside during the creation of the corpus because they do not belong to the first four parts due to a lack of standard and systematic structure of sound production and recognition.

Only the first part is both extensive and systematic in its structure, and only this part is intended to serve as an empirical reference on the basis of which new theses of the acoustic representation of vowel quality are verifiable or falsifiable. The second part aims at a documentation of speech and occurring f_0 ranges and at highlighting the significance of extensive f_0 variation. However, this documentation is not of a systematic structure. (Future language-specific databases will have to address the question of a systematic sample of utterances that could serve as an empirical reference for observable f_0 variation in intelligible speech.) The remaining three parts document experiment-specific sounds and sounds that were recorded in a nonsystematic manner.

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 database 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 20 non-professional and 20 trained and professionally active speakers and singers (hereafter non-professionals and professionals). 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 non-professionals, with no production parameter variations except f_0 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 of Singing

Part 2 of the corpus consists of speech extracts produced by speakers without formal vocal training, by politicians, journalists and TV hosts as well as by professionally trained speakers from the field of the performing arts (above all actors/actresses 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 or Internet content or DVDs/CDs. The extracts document the observable f_0 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 in minimal pairs 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.

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 the next chapter).

Part 5: Miscellaneous

Part 5 of the corpus consists 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 considered by the author to have the ability to satisfactorily produce vowel sounds of sufficient quality for all investigated production parameters, (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 non-professionals), (3) sounds produced by some of the speakers in sVsV context at f_0 levels in a lower frequency range that do not correspond to the standard range of part 1 as well as (4) and duplicates of natural sounds used for specific experiments and related vowel and/or pitch recognition tests. 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 part 5. In addition, for some speakers, glissandi were also recorded.

Published Version 1 of the Corpus

Based on this sample of the work database, the first published version of the Zurich Corpus (Maurer et al., 2018) presented selected sounds of part 1. Since, in many cases, two or multiple recordings were made for a single speaker and a specific configuration of production parameters when creating the database, for publication, a systematic subset of sounds was compiled: If only one sound was recorded for a specific configuration of production parameters, then this sound was selected; if two or more recordings for a specific configuration of production

parameters were made, then the sound with the highest recognition rate, the longest duration and the smallest difference of f_0 intended by the speaker and average f_0 calculated for the radiated sound was selected (according to this order). For non-style productions and each level of vocal effort separately, the sound selection was further limited to an f_0 range for which, for a single speaker in question, all vowels investigated were represented by a sound. (Note that not all speakers were able to produce a complete set of investigated vowels at the very lowest or highest margin of their vocal range. Therefore, f_0 levels with an incomplete set of vowel sounds were excluded.) For productions in style mode, the f_0 range was generally set at the discretion and the style-specific range of the artist.

As a result, a systematic corpus with one sound per speaker and per single production task was created for publication.

The main body of the published version 1 presented c. 33 700 recordings of sounds of all long Standard German vowels, read texts and songs/arias produced by 16 non-professionals (adults and children, gender-balanced) and 24 professionals from the fields of straight theatre, contemporary singing and European classical singing (gender-balanced), with extensive variation of basic production parameters as described above. The side body of the published version 1 presented 830 recordings of sounds of all long Standard German vowels (V context, medium vocal effort) and of read texts produced by 30 native German non-professional reference speakers (see above).

The first published version of the Zurich Corpus thus encompassed c. 34 500 recordings in total, with sound- and speaker-related information, graphic and numerical display of the results of acoustic analysis and of results of the standard listening test. The published corpus was endued with a graphic user interface and additional functionalities (playback, search, export of information and sound download). Also, a Klatt synthesiser as a web application was integrated into the corpus (KlattSyn tool, see next chapter). However, restrictions for the use of the corpus applied since the use of the database was limited to scientific purposes only.

Published Version 2 of the Corpus

The second published version of the Zurich Corpus, presented online in the context of the publication of the treatise Acoustics of the Vowel – Preliminaries (Maurer et al., 2023), consists of the sounds of the first version (with minor corrections) and the following additional selected sounds from parts 2–5 of the entire work database, these sounds being related either to the tables and figures of the treatise mentioned or to the Additions section (see below) of the corpus: Speech extracts, syllables and minimal pairs, filtered sounds, resynthesised and synthesised sounds and sounds of /ɔ/ and of /ə/. In total, at the present moment (date of database publication), c. 37 250 natural sounds and manipulated or artificially produced sounds are presented. (Note that this number will change with database updates as additional sounds will be added, above all sounds related to the Addition section.) Also, the graphic user interface of the corpus was further developed. In addition to the Klatt synthesiser (KlattSyn tool), it now features online tools for sinewave synthesis, harmonic analysis and (re-)synthesis, and sound filtering (SinSyn, HarmSyn and FiltSpec tools, see next chapter).

Furthermore, in this second published version, a new separated section is added and will be updated step by step (see Maurer et al., 2023, entry page, Additions). On the one hand, this section presents topic-specific sound examples selected from the entire sound database that

are often related to this treatise, with the aim of extending the exemplary sound documentation. On the other hand, this section will list and outline in short possible new experiments that allow for an empirical exploration of the phenomena discussed in this treatise.

Concluding, the Zurich Corpus allows for a direct comparison of acoustic characteristics of vowel sounds for intra- and inter-speaker variation of basic production parameters, for vowel resynthesis based on the results of acoustic analysis, for vowel synthesis and for vowel sound filtering. The corpus aims at contributing to a phenomenology of the acoustics of the vowel, that is, building up large-scale, language-specific sound descriptions, addressing all basic variations of production parameters and their possible relevance for vowel quality recognition and voice quality classification.

Table 1 gives an overview of the general structure of the Zurich Corpus, Version 2 (extract of Maurer, 2023, Chapter 1.1).

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

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

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