THE ZURICH CORPUS OF VOWEL AND VOICE QUALITY

VERSION 2

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

ASSISTANT

According to the general structure of the user interface, the **menu** items are shown on the left, the **selected records** are shown on the right, and the top row includes a **player** and indicates the **actual layout** and the **numbers and** (if the "i" is opened) **parameters of the actual sound selection**.

Note, that the full structure appears after selecting sounds.
Note, that JavaScript must be activated, and Cookies must be allowed.

Below, the navigation and the functionality of the menu items are described.

In the Zurich Corpus, open access playback functionality is given for all sounds for which, in the figures and the tables of the treatise Acoustics of the Vowel – Indices (Maurer, 2023, in publication), sound links are provided. For all other sounds, a login is required for the playback functionality (see the handbook of the corpus).

NAVIGATION

The navigation items (top, on the left) allows for a page turning within a layout of presentation. The actual page displayed as well as the number of pages of the selection are indicated in the first row on the top. Typing a page number into the page number field allows for a direct access to this page.

A click on the $\boxed{i}$ in the top row displays the selection parameters of the actual sound selection.

A click on the $\boxed{x}$ in the top row hides the menu.

A click on **>** displays the menu again.

MENU ITEMS

Title page opens the title page of the Zurich Corpus of Vowel and Voice Quality.

Introduction displays a short description of the corpus.

Help includes the present **Assistant**, a help text **How to search** and a help text how to use the **Lightbox**.

Abbreviations lists and details the abbreviations used in the corpus.

Search sounds opens the search form. See the How to search in the Help menu for details on the search parameters.

Statistics includes five menu items: **Categories** opens the list of assigned categories assigned to the sounds of the actual sound selection. **Speakers** opens the list of the speakers of the actual sound selection. **Overview** opens a table presenting the types of production parameters investigated and the corresponding sounds of the actual selection. (For the speakers of the sound selection, see the information on the top of the table.) **Production matrix** opens a table presenting single configurations of production parameters and fundamental frequencies, indicating for each single position the number of sounds of the actual selection. **Breakdown** allows for several listings of database information (for details, see the corresponding popup menu.) – All numerical indications include a link to the related sounds.

Layout offers a selection of six different display layouts for presentation of the actual sound selection.

Sort opens the list of sort parameters offered and indicates the actual parameter selection.

Pictures per page allows for a setting of the number of pictures (single graphical presentations related to the acoustic analysis of a sound) to be displayed within the frame of one page. This option is useful to adapt the presentation to your screen size, to the speed of your internet connection and to your work approach.

Export allows for an export (download) of the record numbers of the actual selection (except all sounds). See the How to search in the Help menu for details on export and import of record number selections.

Lightbox opens a new browser window. Please refer to the separate help text for the lightbox functionality.

On the bottom, the actual selection is indicated, and a link is provided to switch to the entire sample of all records of the corpus.

DISPLAY OF SOUND RECORDS

Below, the standard information in the display of a sound record in corpus is given separately for, firstly, natural sounds in V or sVsV or syllable or minimal pair or word context and, secondly, for texts read or sung or speech extracts on the other hand, except for the first line of the sound legend, which is uniform.

For all sounds, the following indications are shown in the first line of the sound legend: ID number of the speaker, gender (w or m), age (adults = A, children = C), record number of the sound in the Corpus.

For vowel sounds produced in V or sVsV or syllable or minimal pair or word context, the following indications are shown in the second line of the sound legend: Vowel quality intended by the speaker, vowel quality recognised by the majority of the listeners (if a majority was obtained), measured f_0 in Hz or "w" for whispered or "c" for creaky phonation, nationality of the speaker, phonation mode of sound production, vocal effort (medium = med, low = low, high = hgh), and the five individual vowel quality assignments of the five listeners (if tested). – For these sounds, the formant frequencies of LPC analysis (if applied) are given in the third line of the sound legend for a standard parameter setting that accords to the age or gender of the speaker (P4 = standard for children, P5 = standard for women, P6 = standard for men). – In lines 3–6 of the sound legend, all three patterns of formant frequencies of LPC analysis (if measured) for all three parameter settings are given. – If details are displayed by opening the "i" information, details are listed for sound and file information, speaker information and the results of the listening test (if conducted). Furthermore, all three formant patterns of LPC analysis for the three standard parameters applied are given in full, that is, including averaged formant levels and bandwidths. Finally, indications on the sound selection (ranking) and the analysed sound nucleus (sound probe selection) are also listed (for details on the ranking and the determination of the sound nucleus analysed, see the handbook of the Zurich Corpus).

For speech and singing in terms of texts read (tr) or texts sung (ts), the following indications are shown in the second line of the sound legend: intention of production (//text//), nationality of the speaker, phonation mode of sound production, phoneme context (text) and vocal effort (medium = med, low = low, high = hgh, vocal effort variation = var). – If details are displayed via opening the "i" information, details are listed for sound and file information and for speaker information.

Exceptions to the above standard: Exceptions to the above standard occur for duplicates of natural sounds or extracted sound nuclei used for recognition tests whose results are not or not fully related to a single sound record, for some of the resynthesised sounds and for the synthesised sounds. In these cases, some of the standard indications are omitted. For synthesised sounds, also, the ID number of the speaker is replaced by an artificial ID number. Finally, in a few experiments, the display of F-patterns and LPC curves is disabled so as to put forward a direct spectral perspective.

A note on estimated F-patterns as given in figures and tables and in the Zurich Corpus: In general, estimated F-patterns as given in figures and tables correspond to the patterns as given in the Zurich Corpus in terms of default patterns (age- and gender-related default parameters applied in LPC analysis). In some cases, marginal differences occur due to corrections of on- and offsets of sounds and to a recalculation of the patterns when updating the corpus. If not further specified, these differences are ≤ 5 Hz and are neglectable.

Standard of graphic illustration in the Zurich Corpus: For natural vowel sounds produced in V or sVsV or syllable or minimal pair context, graphic representation includes the display of the entire sound wave, the sound nucleus analysed, 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 analysed sound nucleus are overlaid on the spectrum in order to illustrate the correspondence between spectral peaks and calculated formants.

For read texts, songs/arias and speech extracts, graphic representation includes the display of the entire sound wave and the respective f_0 contour, spectrogram and LTAS (0–5.5 and 0–11 kHz).

For manipulated, resynthesised and synthesised sounds, graphic illustration either corresponds to this standard or the illustration is specified in the method section of a given experiment.

Sound pressure level is given in terms of relative levels in dB, adapted for the display of the spectra.

SOUND PLAYBACK

It is imperative to use state of the art headphones to listen to the sounds presented in the Zurich Corpus that feature a playback option; otherwise, sound quality is often significantly impaired. Above all, the use of PC loudspeakers and of other loudspeakers and headphones with nonlinear frequency characteristics will often cause unclear perception and recognition of vowel sounds. In some cases, the recognised vowel quality may indeed be affected as a direct effect of sound distortion.

In all layouts of the Zurich Corpus, sound playback relates to sounds that are normalised to the same rms level 0.2 relative to the maximum. Except is the Layout S: In this layout, three playback options are included related to the normalised sound, its nucleus used for acoustic analysis and the original sound.

SOFTWARE TOOLS

For the implemented software tools KlattSyn, HarmSyn, SinSyn and SpecFilt, please refer to the handbook.