



VOXplot

Acoustic Voice Quality Analysis
Simple | Standardized | Intuitive

Manual

Jörg Mayer

voxplot.lingphon.com

03 | 2024

Disclaimer of warranty

VOXplot is distributed in the hope that it will be useful, but without any warranty; without even the implied warranty of merchantability or fitness for a particular purpose. Please note that **VOXplot** is not a certified medical device and is therefore only recommended for research purposes within the EU. For use outside the EU, the relevant regulations of the respective country apply.

License

VOXplot is open source and released under the GNU General Public License v3.0.

Source code: voxpath.lingphon.com/en/download/

GPL v3.0: www.gnu.org/licenses/gpl-3.0.html

VOXplot uses Praat (v6.0.48) for acoustic analysis. The executable packages which can be downloaded from voxpath.lingphon.com/en/download/ include Praat as an executable binary. Interested parties can find the source code of Praat on GitHub:

Praat source code: github.com/praat/praat

The **VOXplot** manual is licenced under the Creative Commons licence *Attribution-Non-Commercial-NoDerivs 3.0* (CC BY-NC-ND 3.0) which means you are free to share the manual under the following terms: You must give appropriate credit, you may not use the material for commercial purposes, and you may not distribute a modified version of the manual.

CC BY-NC-ND: creativecommons.org/licenses/by-nc-nd/3.0/.

Support



If you use **VOXplot** regularly and would like to support our project, we would appreciate a donation. The QR code will take you to a Paypal page where you can make a one-time donation or an annually recurring donation. **We would like to thank all donors!**

Acknowledgements

I would like to thank Ben Barsties v. Latoszek for his initiative and support, as well as for numerous discussions and invaluable input in the development of **VOXplot**. I also have to thank him sincerely for the contributions and corrections to this manual.

Contents

1	Introduction	5
2	Installation, First Start and Uninstallation	8
2.1	Windows	8
2.2	macOS	8
2.3	Linux	9
2.4	Source-Code	9
2.5	First Start	9
2.6	Uninstallation	10
2.7	Update VOXplot	10
3	The Main Application Window	12
4	The VOXplot Workflow	13
4.1	Step 1: Enter Metadata	13
4.2	Step 2: Voice Samples	14
4.2.1	SNR Evaluation	15
4.2.2	Open Existing Voice Samples	17
4.2.3	Record New Voice Samples	18
4.2.4	Editing Voice Samples	20
4.2.5	Checking and Saving Voice Samples	22
4.2.6	Removing Voice Samples	23
4.3	Step 3: Analysis and Results	24
4.3.1	Output Window	26
4.3.2	Voice Profile	27
4.3.3	Saving Results	29
5	Settings	31
5.1	Default Recording Duration	31
5.2	Default Examiner	31
5.3	Language Settings	32
5.4	Formats	32
5.5	Real-Time Signal Plot	32
5.6	Active Recording Device	33

6	Interpretation of Acoustic Voice Parameters	34
6.1	Preliminary Remarks	34
6.2	Spectral Slope (Slope/Tilt)	35
6.3	Amplitude Difference H1-H2 (H1H2)	36
6.4	Perturbation (Jitter/Shimmer/PSD)	40
6.5	Harmonicity (HNR/HNR-D/GNE/HF Noise)	41
6.6	Voice Breaks	42
6.7	Cepstral Peak (CPPS)	43
6.8	Fundamental Frequency (Pitch/Range)	46
7	Multi-Parametric Voice Quality Indices	48
7.1	Acoustic Voice Quality Index (AVQI)	48
7.2	Acoustic Breathiness Index (ABI)	50
8	Recording: Tips and Recommendations	51
8.1	Hardware	51
8.2	Room Acoustics	53
8.3	Speaking Behavior	54
8.4	Face Masks	54
	References	55

1 Introduction

The analysis of acoustic voice parameters forms an objective-quantitative supplement to conventional, subjective-auditive voice diagnostics. However, although the instrumental overhead is low (a computer is usually available, a microphone suitable for voice analyses is a manageable investment, see suggestions in chapter 8), the integration of acoustic voice analysis into everyday voice therapy has so far been a challenge. In our experience, the biggest hurdles for users of voice analysis software tools are that they are not intuitive to operate and do not offer a robust user interface. Consequently, these tools require a considerable training period and provoke operating errors in the hectic day-to-day practice, which can have a negative impact on the analysis result and thus on the validity of the measurement parameters.

With **VOXplot** we would like to remove these hurdles. We have developed **VOXplot** with the aim of providing a tool for the analysis of acoustic voice parameters that is based on proven and reliable algorithms as well as robust evidence and at the same time uncomplicated and intuitive to use.

Proven and Reliable

VOXplot is developed in close cooperation with scientific advisor Prof. Ben Barsties v. Latoszek. The analysis of a total of 14 acoustic parameters and two multidimensional indices is based on proven Praat algorithms with scientifically based presets. It follows the recommendations for the analysis of acoustic voice quality parameters published in the works on the voice quality indices AVQI and ABI (see section 7).

Simple and Intuitive

VOXplot's user interface is designed to be simple and robust. A complete acoustic analysis of voice quality requires just a few clicks. **VOXplot** is multilingual. The user interface is available in three languages: English, German, Dutch. Valid analysis parameters are currently available for 12 languages: German, English, Dutch, Spanish, Japanese, Korean, Brazilian Portuguese, Malayalam, Kannada, Persian/Farsi, Finnish, and Italian.

The result of a **VOXplot** analysis is a concise voice profile with all examination data and measured values on a single page (see figure 1). In addition to the numerical measurements, the result can be easily and intuitively assessed with the help of a norm-value circle that highlights norm deviations in 6 particularly relevant dimensions. A narrowband spectrogram is also displayed for sustained vowels.

VOXplot calculates and documents acoustic voice parameters based on the analysis of the periodicity of a signal, which can be found in a similar form e.g. in the widely used commercial solution *Multi-Dimensional Voice Program* (MDVP) by KayPentax. In addition, spectral and cepstral parameters (such as *Smoothed Cepstral Peak Prominence* (CPPS) or *Glottal-to-Noise-Excitation Ratio* (GNE)) as well as multiparametric voice quality indices like

AVQI and ABI (see chapter 7) are taken into account, which usually have a greater informative power regarding severity and differential diagnosis of voice disorders [17, 42, 63, 64, 68, 71, 78].

The analysis includes the following 14 voice quality parameters, which are explained in chapter 6; in addition, five pitch parameters are estimated (mean, minimum, maximum, standard deviation and range).:

- ▷ Spectral slope (**Slope**)
- ▷ Spectral tilt (**Tilt**)
- ▷ Amplitude difference between 1st and 2nd harmonics in the spectrum (**H1H2**)
- ▷ Frequency perturbation: **Jitter local (%)** and **Jitter ppq5 (%)**
- ▷ Amplitude perturbation: **Shimmer (%)** and **Shimmer (dB)**
- ▷ Period Standard Deviation (**PSD**)
- ▷ Harmonics-to-Noise Ratio **HNR**)
- ▷ HNR Dejonckere & Lebacq (**HNR-D**)
- ▷ Glottal-to-Noise-Excitation Ratio (**GNE**)
- ▷ High-Frequency Noise (**HF noise**)
- ▷ Smoothed Cepstral Peak Prominence (**CPPS**)
- ▷ Voice breaks

With appropriate voice samples – sustained vowel plus continuous speech – the multiparametric indices *Acoustic Voice Quality Index (AVQI)* and *Acoustic Breathiness Index (ABI)* can also be calculated.

Figure 1 on the next page shows the voice profile of a female speaker with a hoarse voice.

VOXplot – Acoustic Voice Quality Profile

VOXplot v2.0.1

Name:	Ms. Jane Doe	Date:	2023-09-08
ID:	123-456-789-0	Time:	10:33 am
DoB:	1970-01-01	Examiner:	nn
Language:	German (de-DE)	CS recording date:	2021-05-03
Notes:	Entrance examination	SV recording date:	2021-05-03

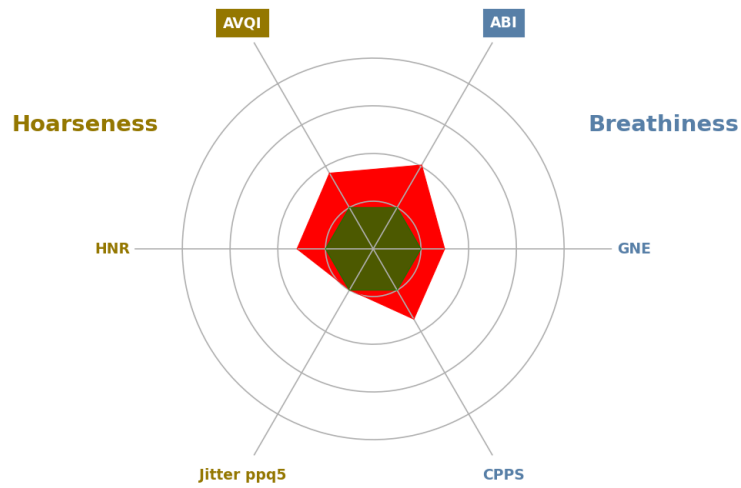
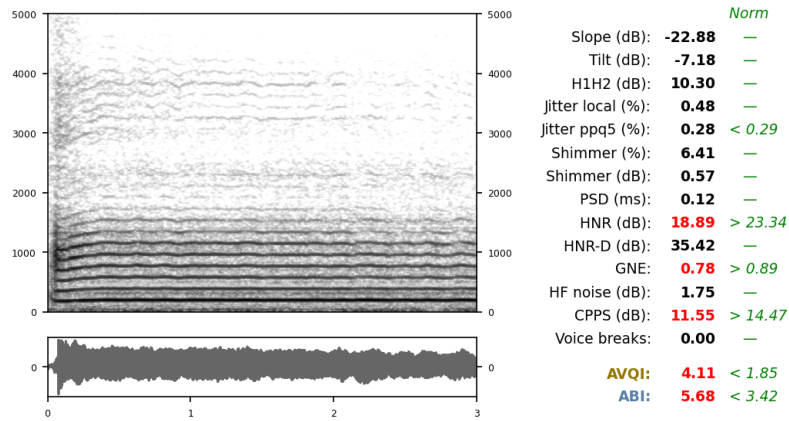


Figure 1: The **VOXplot** voice profile with 2 voice samples: continuous speech (CS) and sustained vowel (SV); female speaker with low-grade hoarseness. Examination data (top), narrow band spectrogram of the SV sample (middle left), measured values [black/red] and norm values [green] (middle right) and norm-value circle with norm deviations (bottom).

2 Installation, First Start and Uninstallation

VOXplot is available as an installation package for Windows (32- and 64-bit), as a standalone executable file for Windows (64-bit only) and Linux and as a macOS app. **VOXplot** is implemented in Python and is open source; you can download the source code and run it in a Python environment. All variants are available for download here:

`voxplot.lingphon.com/en/download/`

2.1 Windows

VOXplot is compatible with Windows versions from version 7 (including Windows 11).

- ↔ Download the installation package (`VOXplot64_setup.exe` for 64-bit systems or `VOXplot32_setup.exe` for 32-bit systems).
- ↔ After the download is complete, double-click the downloaded package to begin the installation process.
- ↔ For a fresh installation, first select whether **VOXplot** should be installed centrally for all users (requires administrator privileges) or only for the current user. This option is not available for an update installation; **VOXplot** will be updated to the existing location.
- ↔ Next, select an installation language (this selection only applies to the installation process; the language selection for **VOXplot** takes place later), confirm the license (GPLv3) and decide whether a desktop icon should be created.
- ↔ At the end of the installation, you can decide whether **VOXplot** should be started immediately, then you can exit the installation program and delete the downloaded installation package.
- ↔ To start **VOXplot** double-click on the desktop icon (if available) or search for the **VOXplot** entry in the start menu.

The procedure described above is the preferred method of installation under Windows. This way, **VOXplot** is well integrated into the system and starts reasonably fast. If you want to avoid system integration and accept a slower startup, you can also download a standalone executable **VOXplot** file (64-bit only). This variant does not need to be installed, but is started directly by double-clicking. Deleting the file will remove **VOXplot** from your computer.

2.2 macOS

Due to a lack of resources, the macOS app can only be tested in a very rudimentary way and problems are to be expected during installation (this is partly because Apple is putting

more and more obstacles in the way of installing software that does not come from its own store). Due to a lack of hardware, we are not able to test the functionality of **VOXplot** on current devices with an ARM chip (Apple Silicon)—but it’s worth a try. In addition, it takes a very long time for the macOS application to launch, and a splash screen to indicate that the launch is underway is unfortunately not supported at this time.

- ↔ Download and unzip `VOXplot.zip`.
- ↔ You can move the included app to any location (e.g. to the desktop).
- ↔ Double-click on the app to start **VOXplot**.
- ↔ If you encounter problems during installation, please consult our website for some tips and tricks.

Deleting the app will remove **VOXplot** from your computer.

2.3 Linux

VOXplot is only available for the `x86_64` architecture.

- ↔ Download the file `VOXplot` and save it to a location of your choice.
- ↔ Mark the file as executable (e.g. in the file manager under *File properties*).
- ↔ Double-click the downloaded file to start **VOXplot** (or run it in a terminal).

Deleting the file will remove **VOXplot** from your computer.

2.4 Source-Code

- ↔ Download and unzip `voxpath_source.zip`.
- ↔ Further information can be found in the README file, which is included in the zip archive.

2.5 First Start

When you start **VOXplot** for the first time, the **VOXplot** settings window opens (Fig. 2). The default language for the user interface is *English*. To switch to another language, select an option from **Interface language** (the other settings are explained in chapter 5). Then click on `Save`, exit **VOXplot** (`Quit` in the main window top right) and restart the program, only then will the new interface language be activated.

You will only see the settings window with default settings after the very first launch. From then on, **VOXplot** saves your settings permanently, even during updates and after a simple uninstall. To completely remove the settings, follow the instructions in the next paragraph.

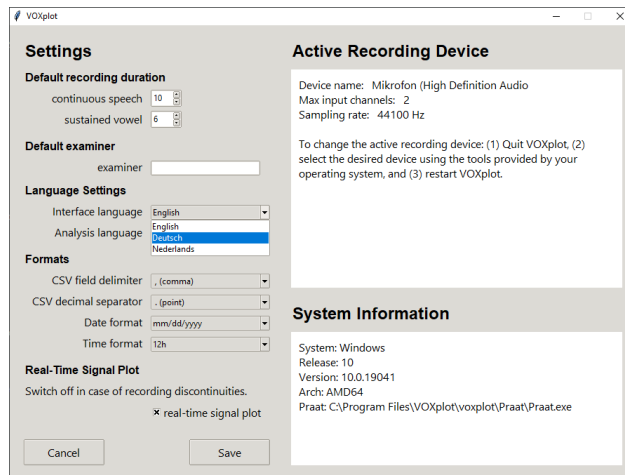


Figure 2: The *VOXplot* settings window after initial installation.

2.6 Uninstallation

Windows

If you have installed **VOXplot** from an installation package, open the Windows settings and then the app overview. Search for **VOXplot** in the list of installed apps and select *Uninstall*. If you have not installed **VOXplot** but have downloaded an executable file, just delete the file. To delete the saved settings after uninstalling or for any other reason, proceed as follows:

- ↪ In the Explorer activate the display of hidden elements
- ↪ Click on your user name in Quick Access and navigate to the folder AppData ▶ Local
- ↪ Delete the folder `voxploit-dir`

macOS & Linux

Simply delete the **VOXplot** app (macOS) or the **VOXplot** executable file (Linux).

To also delete the saved settings after uninstalling or for any other reason, delete the directory `$HOME/.voxploit-dir`

2.7 Update VOXplot

VOXplot is actively developed and updated at irregular intervals in order to eliminate possible errors and/or implement new features. Information on the changes to a new program version can be found here:

voxploit.lingphon.com/en/posts/

You can of course find out whether a new program version is available for download on the homepage, but you can also check within **VOXplot** whether an update is available. To do this, click on **About** in the main window. In the info window you will find the button **Check for updates**. A click on this button starts a request to our server¹ and will then inform you whether you are already using the latest version or whether an update is available. If an update is available, you can download and install the new program version from the download page:

voxplot.lingphon.com/en/download/

¹Privacy Policy Info: Your IP address is transmitted to our server when you make a request. This is technically necessary in order to answer the request. The IP address is stored in the server logs for a maximum of 7 days and then deleted.

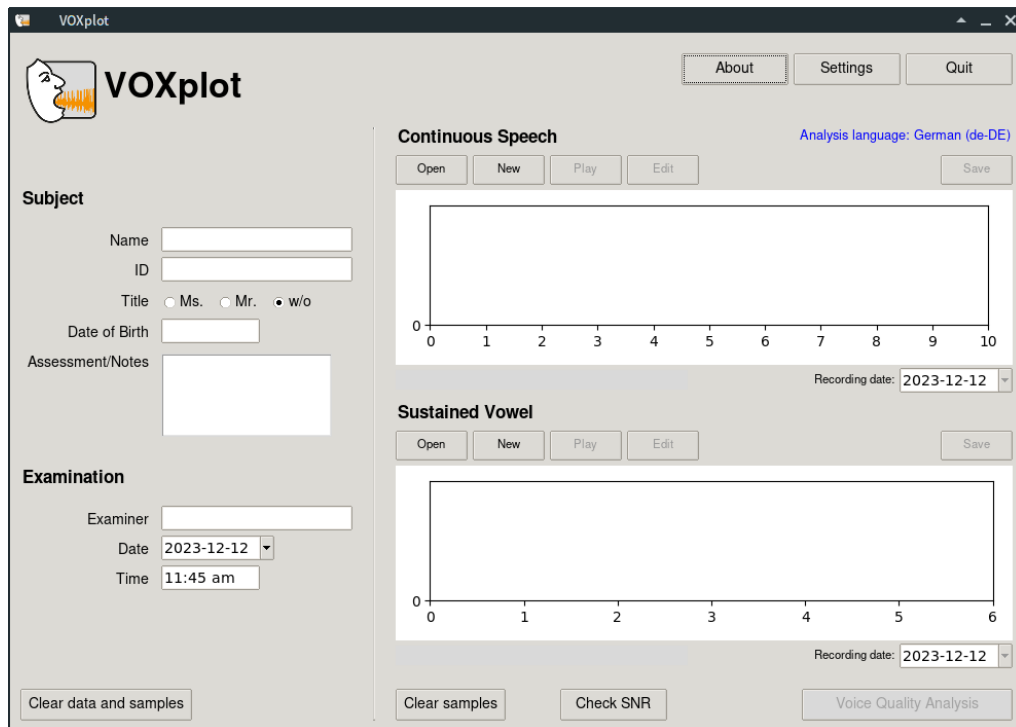


Figure 3: *VOXplot* main application window.

3 The Main Application Window

The main application window of **VOXplot** is divided into three areas (Fig. 3):

- ▷ Three buttons at the top right: **About**, **Settings** and **Quit**
- ▷ On the left, the input fields for metadata
- ▷ On the right, the control elements for two voice samples (continuous speech and sustained vowel)

The **Info** button opens a window with general information about **VOXplot** and a button that checks whether **VOXplot** updates are available (see section 2.7: Update **VOXplot**).

The **Settings** button opens a window in which various default settings can be made and which you already know from section 2.5 (see chapter 5: Settings).

Press **Quit** to exit **VOXplot** immediately and without further prompting.

You will learn about the remaining UI elements in the following chapter.

4 The VOXplot Workflow

The VOXplot workflow comprises three straightforward steps, which are explained in the subsequent sections.

1. Enter meta data (optional)
2. Provide voice samples (open existing samples or record new samples)
3. Start analysis

4.1 Step 1: Enter Metadata

In the main window's left-hand area, you can optionally enter various examination data. Available data appears in the voice quality profile (Fig. 1), but also in the output window (see section 4.3.1) and in the CSV export (see section 4.3.3).

Input fields in the **Subject** section:

<input type="text" value="Namen"/>	:	Name of the person being examined
<input type="text" value="ID"/>	:	some ID (if required)
<input type="text" value="Title"/>	:	Form of address
<input type="text" value="Date of Birth"/>	:	DOB of the person being examined
<input type="text" value="Assessment/Notes"/>	:	Diagnosis, details of therapy or similar.

Information about the **Examination** can be provided in the next section:

<input type="text" value="Examiner"/>	:	Name or initials of the examiner
<input type="text" value="Date"/>	:	Date of examination
<input type="text" value="Time"/>	:	Time of examination

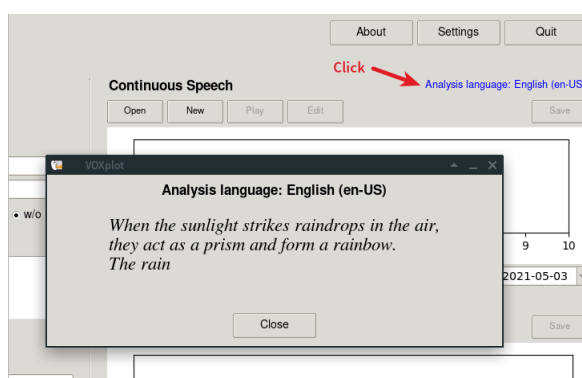
The date and time are automatically set to the current date and time, but can be changed if necessary. If an examiner's name or initials have been specified in the settings (see section 5.2), this field is also filled in automatically, but can also be changed.

4.2 Step 2: Voice Samples

For a complete analysis including the two important indices AVQI and ABI two voice samples are required:

1. Continuous speech (cs)

The text to be read depends on the chosen analysis language (see section 5.3). The analysis language must be chosen according to the native language of the person being examined. The activated analysis language is displayed in the main window; click on it to open a window with the corresponding reading text.



2. Sustained vowel (sv)

A 3 second section from the sustained vowel /a/.

This voice sample specification meets the requirements for calculating AVQI and ABI. (see chapter 7: Multi-Parametric Voice Quality Indices) and only the combination of sustained vowel and continuously spoken text allows a voice evaluation *"that is truly representative of daily speech and voice use patterns"* [69].

Of course, VOXplot can also analyze any other voice samples, but **reliable and comparable results are only obtained with voice samples according to the above specification.**

Existing voice samples can be opened in VOXplot (section 4.2.2). Alternatively, voice samples can be recorded directly in VOXplot (section 4.2.3). Both options can also be combined (voice sample 1 is opened, voice sample 2 is recorded or vice versa).

Attention

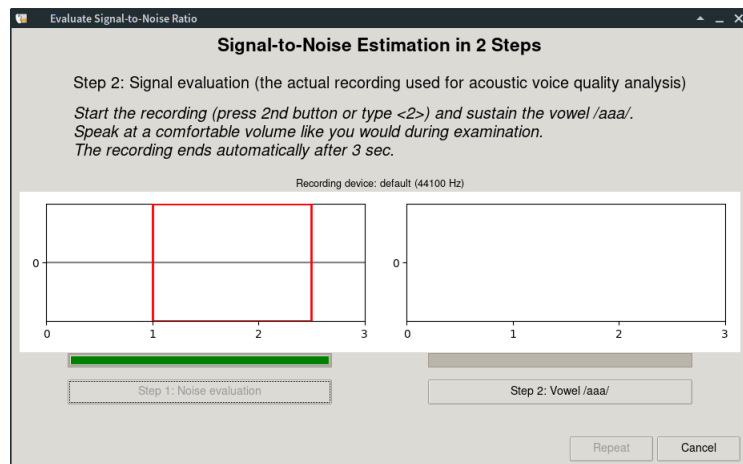
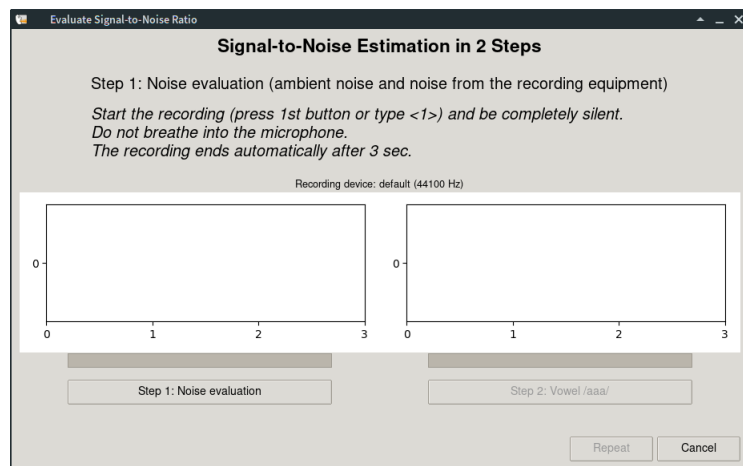
VOXplot only processes audio data in WAV format, other audio formats (such as MP3 files) are not suitable. If you record voice samples outside of VOXplot, make sure to save them as WAV files. Voice samples that are recorded within VOXplot are automatically available in WAV format.

4.2.1 SNR Evaluation

An important criterion for determining whether a voice sample is suitable for acoustic analysis is the signal-to-noise ratio (SNR), which provides information on how well the signal (e.g. the sustained vowel) stands out from the noise floor that accompanies every audio recording. The noise floor is made up of the background noise (which should be largely reduced during recording) and noise components caused by the recording equipment (microphone, amplifier, converter, etc.).

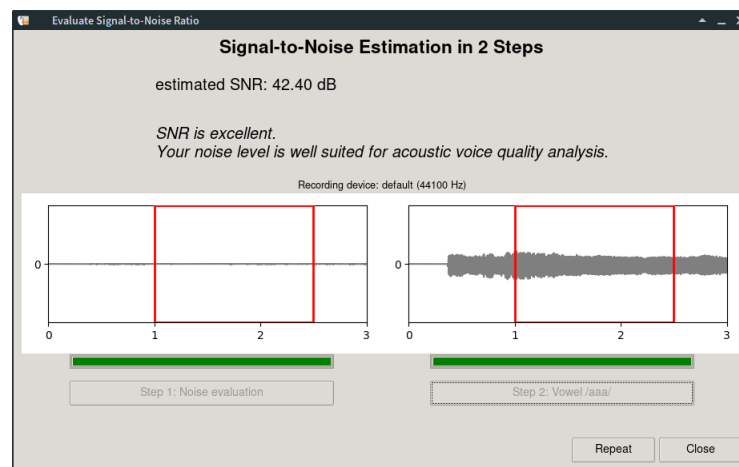
Since version 2.0, VOXplot offers the option of evaluating SNR for a specific recording setup. The recording setup includes the room in which the recording takes place, the distance between the microphone and the speaker as well as the recording devices used and their settings (e.g. the microphone gain). To evaluate a recording setup, set everything up as for a routine examination, then start VOXplot and click the button **Check SNR** (at the bottom of the main window). SNR evaluation consists of two simple steps:

1. First, only the noise floor should be recorded, i.e. you start a recording by clicking the **Step 1: Noise evaluation** and are completely silent.



Recording ends automatically after 3 seconds and the recorded signal appears. Only the section marked in red is taken into account for SNR evaluation. If there is unwanted noise in this section (e.g. a rash from a slamming door or similar), you should cancel the procedure and repeat it.

2. The second step consists of recording a sustained vowel under identical conditions as during an examination. Start the recording by clicking **Step 2: Vowel /aaa/** and produce an /aaa/ (or have a patient produce an /aaa/). The recording ends automatically after 3 seconds and the signal is displayed. The result of the SNR measurement and an assessment of the SNR with regard to suitability for acoustic voice analysis appear above the signal display (too low / acceptable / excellent).



The signal-to-noise ratio is specified in dB. An SNR ≥ 30 dB is required for a valid and reliable acoustic voice analysis, highest accuracy is achieved with an SNR > 42 dB [5, 34, 75].

In order to improve the signal-to-noise ratio, it is necessary to check the entire recording setup:

- ▷ Reduce background noise (computer fans, building ventilation/air conditioning systems, whirring power supply units, etc.).
- ▷ If no headset microphone is used: Reduce the distance between the speaker and the microphone.
- ▷ Possibly adjust the microphone gain.
- ▷ Use other devices (microphone, microphone amplifier, etc.).

Attention

The SNR evaluation is only valid for a specific recording setup! The evaluation should be repeated if the setup is changed.

4.2.2 Open Existing Voice Samples

1. Depending on which voice sample you want to open in VOXplot click the **Open** button in the corresponding section of the main application window (**Continuous speech** or **Sustained vowel**).



2. In the file selection dialog, select the corresponding voice sample (WAV file).
- 3a. **Continuous speechText (cs)**
If it is a cs voice sample, the file is loaded into the cs editor to extract the relevant section for analysis from a longer recording → *CS Editor* (p. 20).
 - 3b. **Sustained vowel (sv)**
An sv voice sample that corresponds to the required duration (3 sec.) is transferred directly to the main window.

If the sv voice sample is shorter than 3 seconds, a warning appears that the sv sample is not suitable for a standardized AVQI or ABI analysis, but it is still loaded and can be analysed. Section 4.3 explains which analyses are performed in this case and which are not.

If the sv voice sample is longer than 3 seconds, it is loaded into the sv editor and can be shortened easily by moving a selection box → *SV Editor* (p. 21).

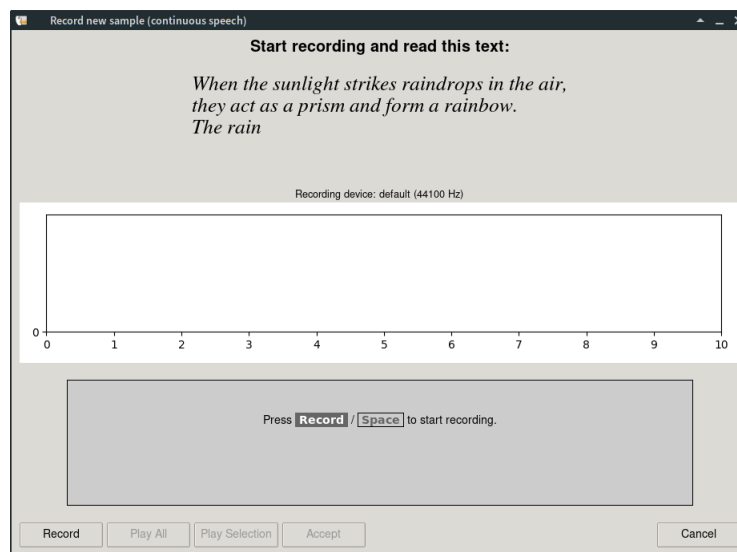
4.2.3 Record New Voice Samples

1. Depending on which voice sample you want to record, click the **New** button in the corresponding section of the main application window (**Continuous speech** or **Sustained vowel**).



2a. Continuous speech (cs)

The cs record window opens:



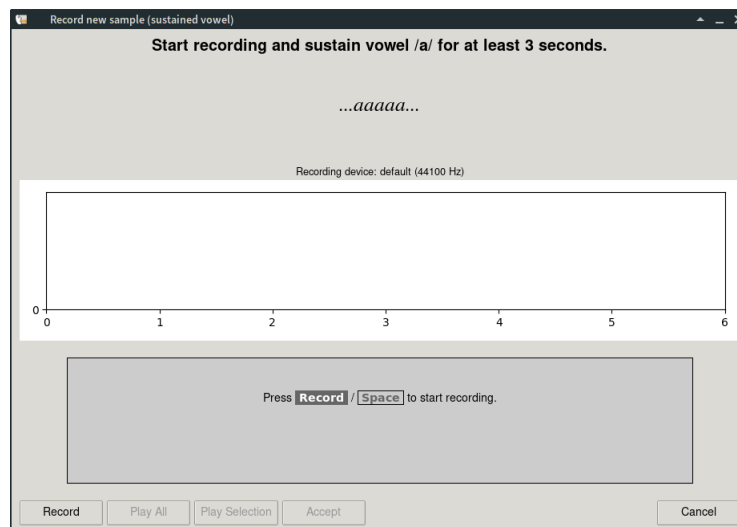
Check above the signal display whether the correct recording device is selected (see section 5.6). Instruct the person being examined to read the displayed text. Click on **Record** or press **Space** to start the recording. During the recording, the signal display is continuously updated; this can be deactivated on less powerful computers (see section 5.5).

When the voice sample is complete, click on **Stop** or press **Space** again. Otherwise, the recording ends automatically after the preset recording time has elapsed (see section 5.1).

Once the recording is complete, the signal is opened in the cs editor and the desired section can be selected and transferred to the main window → **CS Editor** (p. 20).

2b. Sustained vowel (sv)

The sv record window opens:



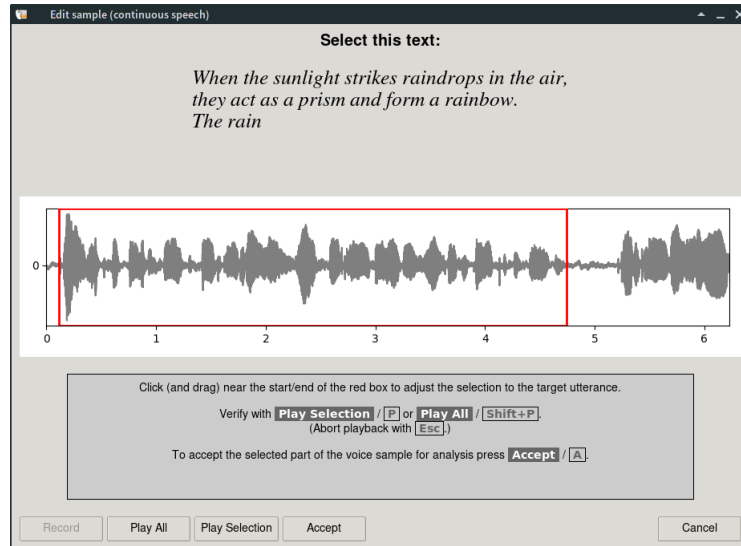
Check above the signal display whether the correct recording device is selected (see section 5.6). Instruct the person being examined to sustain the vowel /a/. Click on **Record** or press **Space** to start the recording. During the recording, the signal display is continuously updated; this can be deactivated on less powerful computers (see section 5.5).

When the voice sample is complete, click on **Stop** or press **Space** again. Otherwise, the recording ends automatically after the preset recording time has elapsed (see section 5.1).

Once the recording is complete, the signal is opened in the sv editor and the desired vowel segment can be selected and transferred to the main window → *SV Editor* (p. 21).

4.2.4 Editing Voice Samples

CS Editor The cs editor is used to trim cs voice samples.



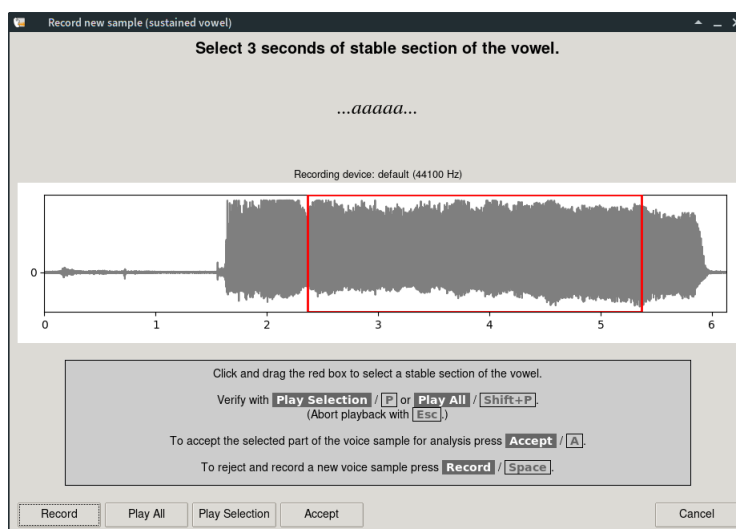
Initially, the signal that is loaded into the cs editor is fully selected, i.e. the red selection box covers the entire signal. The start and end of the selection box can now be moved using the mouse (click—drag—release) so that only the desired signal section is selected.

Ensure that any unwanted noise is outside of the selected area. Check the selection with **Play Selection**; the entire recording can be played back with **Play All**. If you agree with the selection, the marked section can be transferred to the main window with **Accept**.

Keyboard shortcuts:

Button	Key(s)
Play Selection	p
Play All	Shift + p
Abort playback	Esc
Accept	a
Cancel	Esc
Only if the recording function is active:	
Record	Space

SV Editor The cs editor is used to trim sv voice samples.



By moving the red selection box with the mouse, a 3-second section in a stable phase of the vowel can be selected. The size of the selection box is fixed at 3 seconds and cannot be changed, i.e. the box can only be moved to the right or left.²

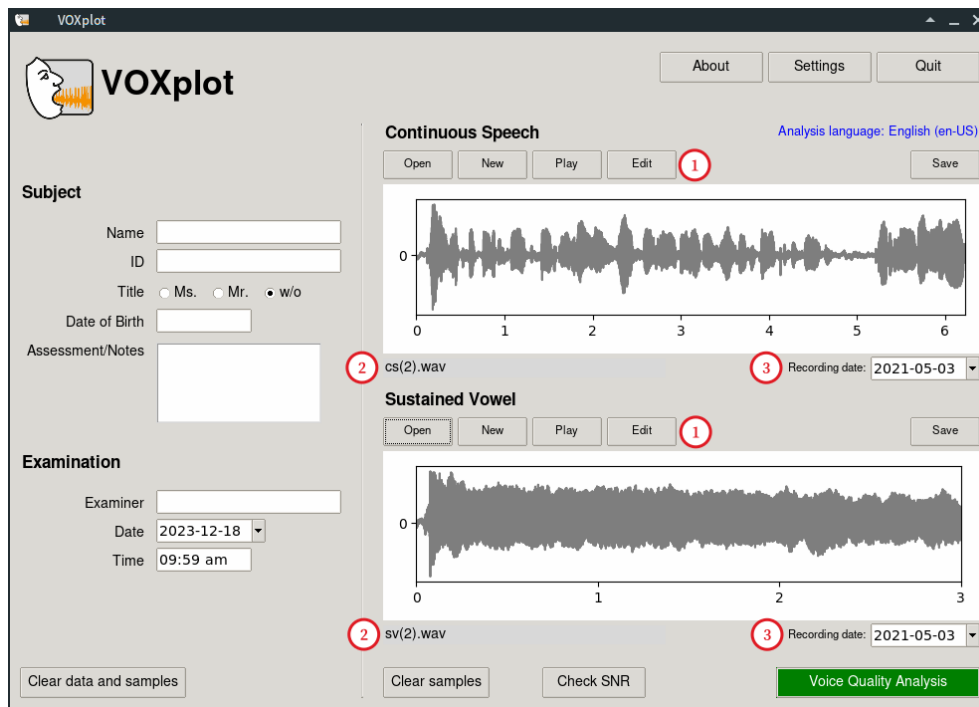
The play buttons **Play selection** and **Play all** are available to check the selection. The selected section is transferred to the main window with **Accept**. Keyboard shortcuts are equivalent to the cs editor.

Record Button If a voice sample is recorded directly within VOXplot the cs or sv editor opens automatically after the recording is finished. In this case, the **Recording** button is also active in the respective editor. If the voice sample turns out to be unsuitable, this button can be used to start a new recording directly from the editor. In all other cases, the recording button is deactivated.

Cancel Button Click **Cancel** to close the respective editor window. If you have jumped directly from a recording to the editor (see section 4.2.3), the recorded voice sample will be discarded. However, if the editor has opened after opening a voice sample (see section 4.2.2) or by clicking on **Edit** in the main window (see section 4.2.5), only the selection is discarded and the voice sample is retained.

²If you want to analyze a vowel sample with a different duration, you must record and trim it outside of VOXplot, then open it in VOXplot and close the sv editor with **Cancel**. The sample is then available in the main window for analysis.

4.2.5 Checking and Saving Voice Samples



1 Buttons

If a voice sample has been opened or recorded, three additional buttons are activated in the section of the respective voice sample: **Play**, **Edit** and **Save**.

Play plays the respective voice sample, **Edit** opens the cs or sv editor and with **Save** the voice sample can be saved permanently to your local hard disk.

Attention

Recordings that are created directly in VOXplot must be saved with the **Save** button if they are to be retained permanently.

2 Filenames

If a voice sample has been opened, the file name is displayed below the signal display (in the example: *cs(2).wav* below the cs voice sample). If a voice sample was recorded within VOXplot and has not yet been saved, *unsaved data* is displayed instead; as soon as the sample has been saved, the file name appears.

③ Recording Date

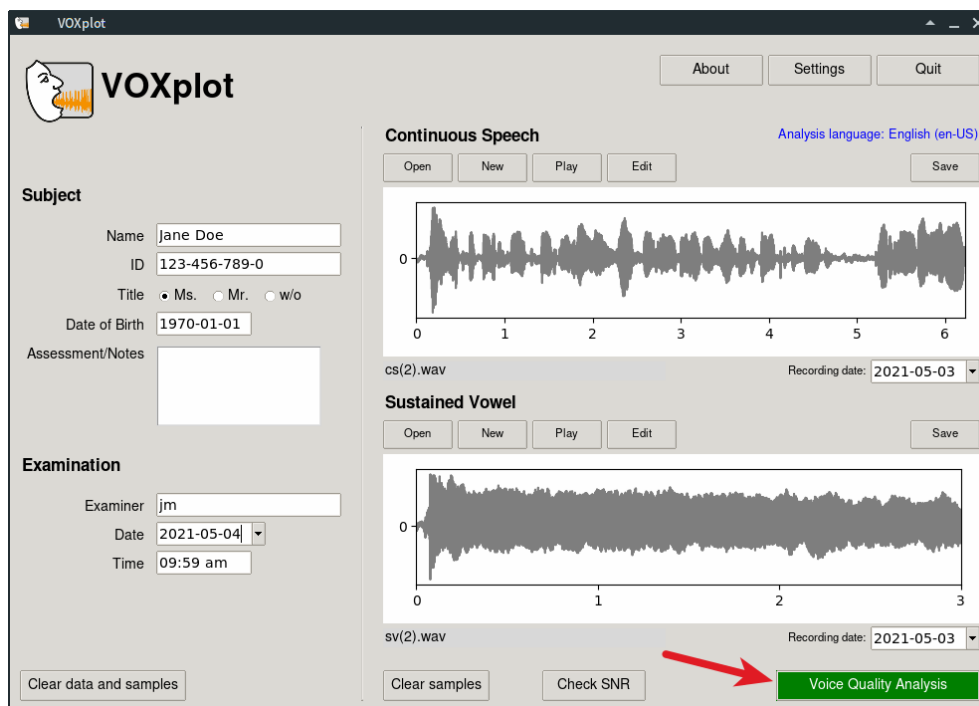
The recording date of each voice sample appears automatically. For samples that were already available on the hard disk and were opened in **VOXplot** the creation date of the file is used; for samples that were recorded within **VOXplot** the current date is used. If necessary—e.g. if the creation date of the file does not match the actual recording date—the recording date can be changed for both voice samples. The recording dates of both voice samples appear in the voice profile (Fig. 1, top right).

4.2.6 Removing Voice Samples

Opening (**Open**) or recording (**New**) a voice sample always overwrites existing voice samples. Explicit deletion of the samples is therefore not absolutely necessary, but possible: The button **Clear samples** removes **both** voice samples from **VOXplot**.

In addition to the voice samples, the button **Clear data and samples** also removes all meta data from the form in the left-hand area and updates the date and time; the button has the same effect as restarting **VOXplot**.

4.3 Step 3: Analysis and Results



As soon as at least one voice sample is available, the green button **Voice Quality Analysis** is activated. If both a cs voice sample and a sv voice sample with a duration of 3 seconds are available, a complete analysis is started with this button. If there is only one voice sample and/or the sv sample is too short or too long, a warning is displayed. The analysis can then be canceled to correct the voice samples or an incomplete analysis can be started. The following overview shows which analyses are carried out, depending on the available voice samples:

Voice samples		Analyses			
CS	SV (duration/sec.)	num. parameters	AVQI	ABI	voice profile
✓	✓ (= 3)	✓	✓	✓	✓
✓	✓ (≠ 3)	✓	-	-	-
✓	-	✓	-	-	-
-	✓ (= 3)	✓	-	-	✓
-	✓ (≠ 3)	✓	-	-	-

Attention

VOXplot only checks the vowel duration. It is not possible to check whether the cs voice sample represents the correct section of text—it is the examiner's responsibility to ensure that the text passage specified by the analysis language setting is actually included in the analysis.

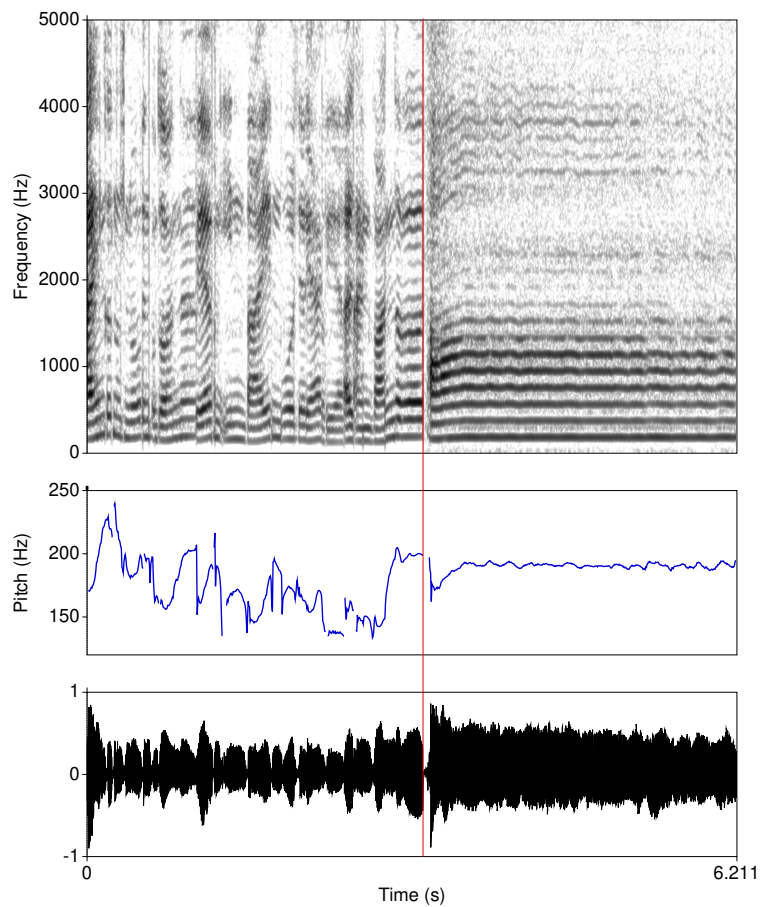


Figure 4: *MX Voice Sample; oscillogram (bottom), fundamental frequency response (middle) and narrowband spectrogram (top). The vertical red line marks the boundary between the phonated parts of the cs voice sample (left) and the sv voice sample (right). The duration of the cs voice sample reduced to phonated parts naturally varies; in the example shown it is 3.211 sec.; combined with the sv voice sample (duration: 3 seconds), this results in a total duration of the mx voice sample of 6.211 sec.*

In a complete analysis, **VOX**plot automatically generates a third voice sample (*mx*) on the basis of the two existing voice samples (*cs* and *sv*). The *mx* sample consists of the complete *sv* sample and the phonated parts of the *cs* sample. The algorithm for recognizing and extracting the phonated parts from the *cs* voice sample is described in [81] resp. [69].³ Ideally, the *mx* voice sample is a fully phonated combination of *cs* and *sv* voice samples, which is used in particular for AVQI and ABI analysis (Fig. 4).

³**VOX**plot largely adopts the Praat implementation of the algorithm by Paul Corthals (University of Ghent, Belgium).

4.3.1 Output Window

Once the analysis has been started, a results output window appears, which gradually fills with content. Depending on the performance of the computer, the analysis may take some time—the green bar at the top of the window shows the progress:

Parameter	Norm	SV	CS	MX
Slope (dB)		-22.88	-22.91	-22.89
Tilt (dB)		-7.18	-9.20	-8.95
HF Noise (dB)		1.75	1.77	1.77
HNRd (dB)		35.42	26.26	31.28
HIH2 (dB)		10.30	4.89	7.76
CPPS (dB)	> 14.47	11.55		

Once the analysis is complete, the meta data is displayed at the top (if available), followed by a table of the numerical results:

Parameter	Norm	SV	CS	MX
Slope (dB)		-22.88	-22.91	-22.89
Tilt (dB)		-7.18	-9.20	-8.95
HF Noise (dB)		1.75	1.77	1.77
HNRd (dB)		35.42	26.26	31.28
HIH2 (dB)		10.30	4.89	7.76
CPPS (dB)	> 14.47	11.55	10.52	10.99
Jitter local (%)		0.48	2.00	1.19
Jitter p95 (%)	< 0.29	0.28	1.06	0.61
Shimmer (%)		6.41	8.52	7.15
Shimmer (dB)		0.57	0.84	0.69
HNR (dB)	> 23.34	18.89	15.20	17.02
PSD (ms)		0.12	0.76	0.58
Voice Breaks		0.00	-	-
GNE	> 0.89	0.78	0.87	0.83
Pitch (mean, Hz)		189.82	175.07	-
Pitch (min, Hz)		162.11	124.76	-
Pitch (max, Hz)		194.88	469.04	-
Pitch (sd, Hz)		3.94	30.42	-
Range (st)		3.19	22.93	-
AVQI	< 1.17			4.11
ABI	< 2.35			5.68

The first column lists the parameters (14 acoustic voice parameters (see p. 6), 5 pitch parameters and 2 indices). In the second column, norm values are listed in green (if available). The norm values of the individual parameters apply to a sustained vowel lasting 3 seconds; the AVQI and ABI norm values are language-specific and apply to the combined voice sample (mx). The third and fourth columns show the results for the sv and cs voice sample. The

fifth column shows the results for the combined voice sample (mx). Results of the sv voice sample that are outside the normal range are displayed in red (if norm values are available), as are AVQI and ABI values above the normal range.

In the case of non-standard voice samples, the output contains incomplete results. The following example shows results of an sv sample with a duration of less than 3 seconds; the cs sample was missing:

Parameter	Norm	SV	CS	MX
Slope (dB)	-	-24.82	-	-
Tilt (dB)	-	-6.54	-	-
HF Noise (dB)	-	1.79	-	-
HNRd (dB)	-	35.51	-	-
HIH2 (dB)	-	7.98	-	-
CPPS (dB)	-	11.66	-	-
Jitter local (%)	-	0.40	-	-
Jitter ppq5 (%)	-	0.24	-	-
Shimmer (%)	-	6.26	-	-
Shimmer (dB)	-	0.55	-	-
HNR (dB)	-	19.36	-	-
PSD (ms)	-	0.04	-	-
Voice Breaks	-	0.00	-	-
GNE	-	0.79	-	-
Pitch (mean, Hz)	-	190.84	-	-
Pitch (min, Hz)	-	186.30	-	-
Pitch (max, Hz)	-	194.82	-	-
Pitch (sd, Hz)	-	1.42	-	-
Range (st)	-	0.77	-	-
AVQI	-	-	-	-
ABI	-	-	-	-

(SV was too short; analysis incomplete.)

In this case, all numerical parameters of the sv voice sample are calculated (3rd column), but no norm values are given (2nd column), as these only apply to vowel samples with a duration of 3 seconds. AVQI and ABI are not calculated because the sv sample is too short and the cs sample is missing.

4.3.2 Voice Profile

If the voice samples meet the specifications outlined in section 4.2, a graphical voice profile will be generated in the same window along with the textual output (Fig. 5). The voice profile is displayed concisely on a single page:

1. **Metadata:** All the data previously entered in the form on the left-hand side of the main window and the recording date of the two voice samples.
2. **Narrowband spectrogram** of the sustained vowel for spectrographic voice classification ([6, 91]).
3. **Numerical results** of the sustained vowel as well as AVQI and ABI index; values within normal range in black color, deviating values in red color; to the right in green color the norm values, if available.

- Norm value circle**, illustrating the analysis results of the sv voice sample in 4 relevant dimensions ([18], see also section 6.1) as well as the two indices AVQI and ABI, which are calculated on the basis of the combined mx voice sample, using an intuitive traffic light system. The middle green circle represents the normal range. Deviations are indicated in red; the further a value deviates from the norm in one dimension, the further out the value is plotted. If the deviation falls outside the displayable range, this strong anomaly in voice quality is marked with red arrows (Fig. 6).

Even if only a sv voice sample with a duration of 3 seconds is available, a graphical voice profile is generated. However, due to the missing cs voice sample, the AVQI and ABI indices cannot be calculated. Therefore, in this case, only the individual parameters are taken into account.

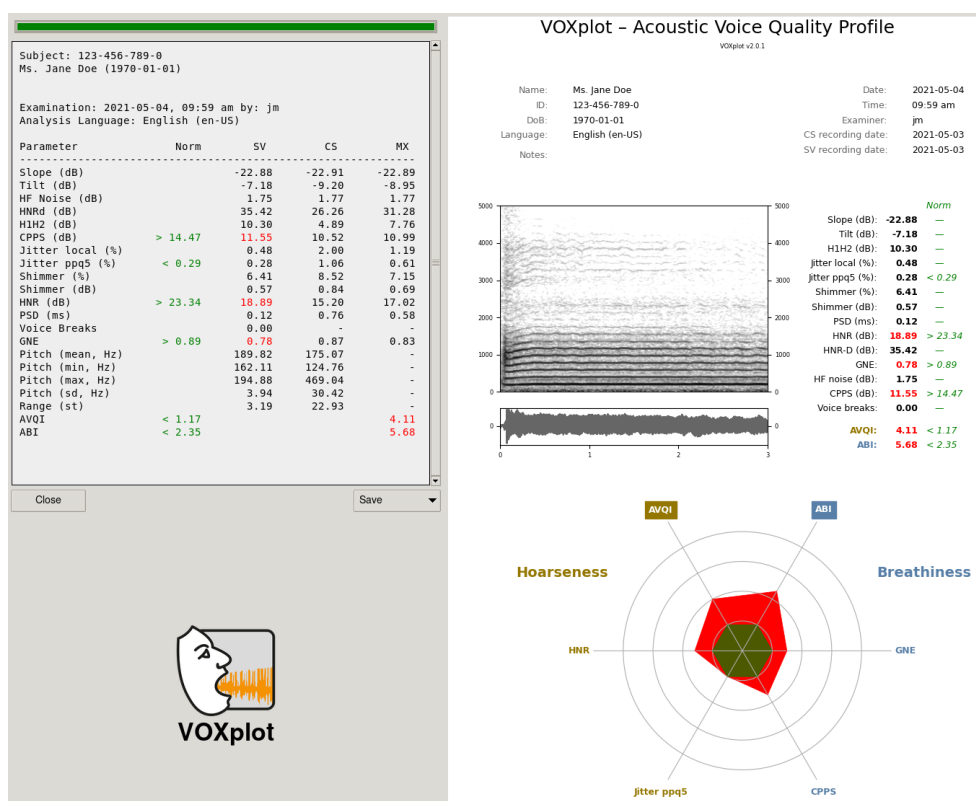


Figure 5: Output of results (top left) and graphical VOXplot voice profile (right).

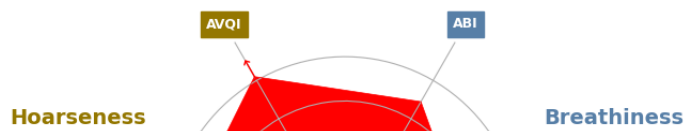


Figure 6: Deviations from the norm outside the displayable range (in this example: AVQI value). This indicates a particularly strong anomaly in voice quality with regard to this parameter.

4.3.3 Saving Results

The graphic voice profile can be saved both as a PDF file and as an image file in PNG or JPG format for archiving, forwarding or printing. The **Save** menu is located below the results.



If the analysis results will be processed in another format, such as Excel, Calc, R, SPSS, or GNU PSP, the meta data and numerical results can be saved in a universal CSV format. This can be done by creating a new CSV file or appending the data to an existing one. The desired CSV format (field separator and decimal point) should be checked beforehand in the **VOXplot** settings and specified if necessary (see section 5.4).

The option **CSV New** creates a new CSV file with 2 lines: The first line contains the header (column titles), the second line contains the exported data. The option **CSV Append** can be used to select an existing CSV file; the exported data is then added to this file in a new line (without header). Attention: The existing file should already have a header in the first line and have the same CSV format as the exported data.

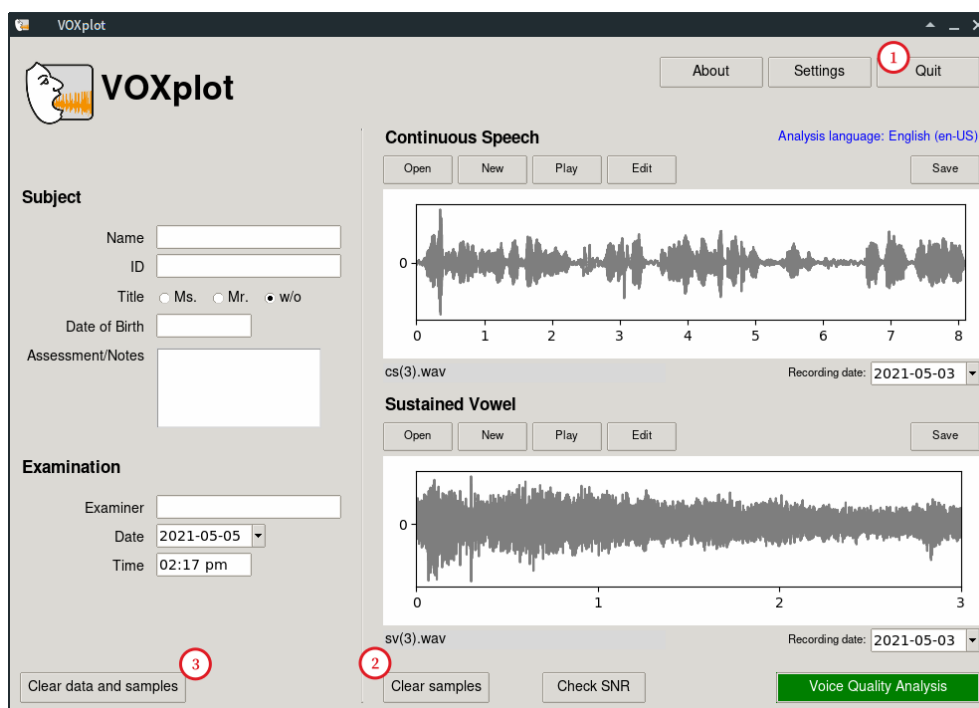
A total of 70 fields are exported: all metadata as well as all numerical analysis results. The following figure shows an excerpt (9 columns) from the resulting table after importing the CSV file into LibreOffice Calc:

The screenshot shows the LibreOffice Calc interface with a CSV file imported. The data is as follows:

	sv_hfno	sv_hnrd	sv_h1h2	cs_hfno	cs_hnrd	cs_h1h2	mx_hfno	mx_hnrd	mx_h1h2
1									
2	1.1721245032128047	16.38375426227333	1.50885678332246	1.501754769333264	19.62434236506296	6.956582978592916	1.3304043887617845	18.703657707670008	5.206375606689481
3									
4									
5									

The excerpt shows the 3 parameters **hfno** (high-frequency noise components), **hnrd** (HNR according to Dejonckere & Lebacq) and **h1h2** (amplitude difference between 1. and 2nd harmonics), each for the sustained vowel (**sv_**), continuous speech (**cs_**) and the combined sample (**mx_**).

After saving the results, the output window can be closed with the **Close** button. You now have a choice in the main window:



① Quit VOXplot

Attention

Voice samples that were recorded within VOXplot and have not yet been saved will be lost.

② Remove Samples

If you would like to open or record further voice samples from the person being examined, you can delete the existing voice samples with **Clear samples**. The metadata in the left-hand area will be retained (and can be edited if necessary).

② Remove Samples and Metadata

If another person is to be examined, **Clear data and samples** deletes not only existing voice samples but also the metadata; examination date and time are updated. This is equivalent to the status of VOXplot after a restart.

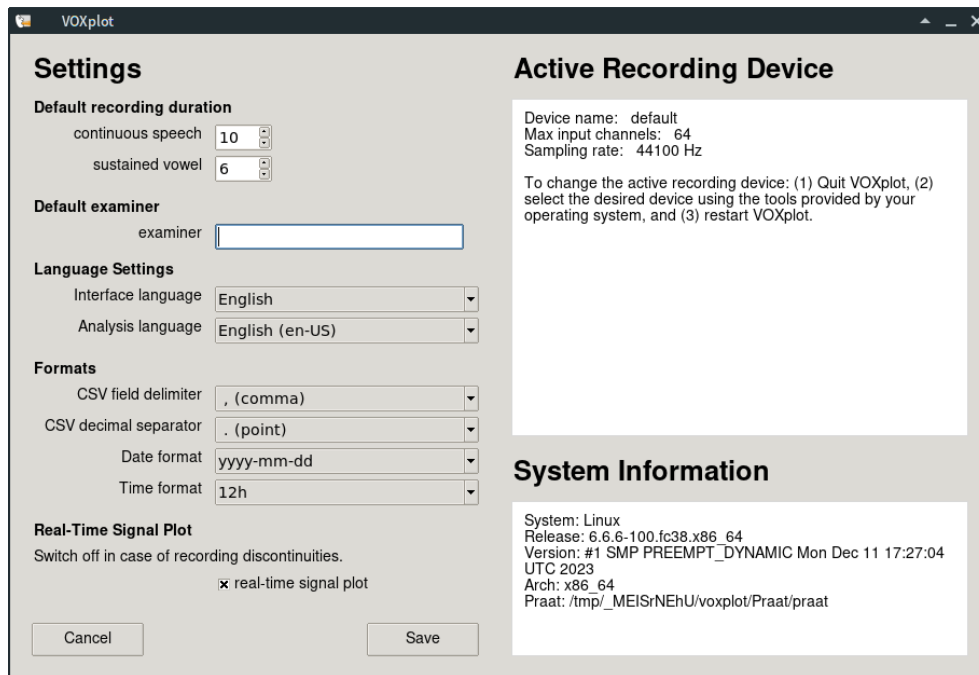


Figure 7: *VOXplotSettings*.

5 Settings

To open the settings, click on **Settings** in the main application window. The settings window shows information about the active recording device and the computer system on the right. In the left-hand window area, you can set preferences that control the behavior of **VOXplot**. These preferences are saved on your computer in your user account and will be retained after closing **VOXplot**.

5.1 Default Recording Duration

The recording duration (in seconds) can be preset separately for continuous speech (cs) and sustained vowels (sv). After the recording duration set here, the respective recording stops automatically (if it has not been stopped manually beforehand). The maximum value for cs recording is 30 seconds, for sv recording 15 seconds. The default settings are 10 and 6 seconds respectively.

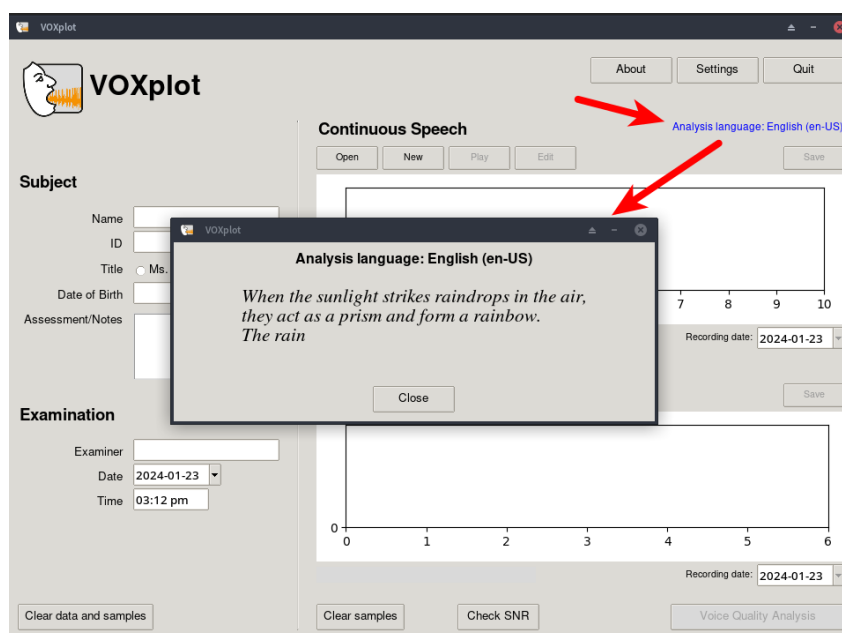
5.2 Default Examiner

If examinations are predominantly carried out by the same person, a name or initials can be entered here. The field **Examiner** in the main window is then automatically filled with this name/initials.

5.3 Language Settings

The **interface language** affects the labels of buttons and other elements of the user interface as well as some labels in the voice profile (see Fig. 1). This setting can be freely selected according to the examiner's preference.

The **analysis language** must be selected solely based on the language being analyzed. The voice of a German-speaking person must therefore be analyzed with the setting *German*, the voice of a Spanish-speaking person with the setting *Spanish*, etc. The selected analysis language is displayed prominently in the main window; clicking on the language opens a window with the reading text specified for this language:



5.4 Formats

VOXplot can export numerical analysis results in universal CSV format (see section 4.3.3). The first two settings in the *Formats* section can be used to specify which character should be used to separate the fields in a CSV data set (default: comma) and which decimal character should be used for numerical data (default: point). Depending on how the CSV data will be processed, other values may be useful.

Date and time format settings affect how the date and time are displayed in the and fields in the main window, as well as the output in the voice profile.

5.5 Real-Time Signal Plot

By default, the signal progression is displayed in (almost) real time during a recording. However, on less powerful computers, this may overload the system and cause dropouts in

the recording. If you notice such dropouts, you should deactivate the real-time signal plot [here](#).

5.6 Active Recording Device

Several recording devices can be connected to a computer at the same time, but normally only one of them is actively used for recording. For example, you can connect a high-quality external microphone to a laptop via USB in addition to the built-in microphone (which should not be used for voice analysis). Which of the connected devices is used for recording can be selected in the operating system (Windows, macOS, Linux) using the built-in audio mixer tools. **VOXplot** recognizes the active recording device and displays it in the corresponding section together with the number of channels and the set sampling rate. If the device displayed is not the desired device, close the settings window and exit **VOXplot**. Then use the operating system tools to select the desired device and restart **VOXplot**.

6 Interpretation of Acoustic Voice Parameters

6.1 Preliminary Remarks

The calculation of acoustic voice parameters is influenced by various factors, which can make it difficult to interpret and compare values. These factors include the type of microphone used, other recording hardware, the acoustics of the room, the selected sampling rate, and most significantly, the algorithms used, for example to detect voiced parts in the acoustic signal or determine jitter [5, 26, 33–36, 82, 93, 96, 97].

For this reason, norm values for certain parameters should be treated with a certain degree of caution. Even the use of a different microphone or a different sampling rate can reduce the significance of such norm values. Norm values become practically useless if the corresponding parameters are calculated using different algorithms. This applies, for example, to **VOXplot/Praat** in comparison with the commercial MDVP [2, 24, 70, 80]. The underlying algorithms differ considerably between the two programs. As a result, the MDVP norm values can only serve as a very rough guide, if at all, when interpreting a **VOXplot** voice analysis, and vice versa. An explanation of how and why the algorithms differ between **VOXplot/Praat** and MDVP, and the consequences for values calculated with Praat compared to MDVP values, can be found in the Praat manual.

VOXplot provides some of its own norm values. The cut-off values for the multi-parametric indices AVQI and ABI are language-specific and will only be included in **VOXplot** if corresponding scientifically sound, language-specific studies are available [1, 9, 15, 17, 32, 38–40, 50, 51, 53, 55, 57, 58]. In addition, since **VOXplot** v2.0, only norm values are given for four individual acoustic parameters of the (non-language-specific) vowel voice sample: HNR, Jitter ppq5, CPPS and GNE. As was shown in a recent study based on 218 voice samples, of the 13 parameters examined⁴ only these four parameters can be validly associated with the perceptual assessment of hoarseness or breathiness [18]. Criteria for assigning a parameter to one of the two voice abnormalities were: (a) a high correlation ($Correlation > 0.70$ and $A_{ROC} > 0.70$) with hoarseness or breathiness and (b) a significant difference in correlation or ROC statistics with regard to hoarseness and breathiness. In other words, only those parameters that were strongly associated with one voice abnormality and significantly less associated with the other voice abnormality met the inclusion criteria. For hoarseness, this applies to the two parameters *Harmonics-to-Noise Ratio* (**HNR**) and the pitch perturbation quotient with a smoothing factor of 5 periods (**Jitter ppq5**). Associated with breathiness are the *Smoothed Cepstral Peak Prominence* (**CPPS**) and the *Glottal-to-Noise Excitation Ratio* (**GNE**).

⁴The 13 parameters are the **VOXplot** parameters which are explained in the following sections, with the exception of the unspecific parameter *Voice breaks*.

Attention

Norm values specified in this manual or in the **VOX**plot results output are not transferable. They are only valid in conjunction with test results obtained with **VOX**plot.

The problem of norm values is alleviated if **VOX**plot is used for follow-up examinations, e.g. as part of the documentation or evaluation of voice therapy, and if the recommendations for recording settings formulated in chapter 8 are followed. In this case, the acoustic voice parameters of a person are gathered at different times and compared to each other instead of to norm values. In this way, positive or negative alterations of the voice can be objectively documented. Deciding whether a parameter has changed for better or worse is usually straightforward. However, such comparisons are only reliable if the recording setup (microphone, microphone distance, other recording hardware, environment, etc.) and the recording procedure are kept as identical as possible across the various examinations.

In general, parameters that require an analysis of the periodicity of the time signal before they can be calculated (e.g. jitter, shimmer or PSD) are more problematic than spectral parameters (e.g. slope, tilt or CPPS). There are many algorithms that can be used to analyze periodicity, but all algorithms are fallible, and the number of errors increases—up to complete failure—the more problematic the signal. In this respect, all signals that are not generated with a modal voice must be considered problematic, i.e. exactly those signals that are often of particular interest in voice analysis.

This problem does not apply to spectral parameters because they are calculated after the time signal has been transformed to the frequency domain using spectral analysis (Fourier transformation). In this respect, spectral parameters are more robust than, for example, perturbation parameters. Another advantage is that, in addition to sustained vowels, natural, continuous utterances can also be analyzed in a meaningful way. This is made possible by the use of a *Long-Term Average Spectrum* (LTAS). In this analysis method, the spectral characteristics of a signal are integrated over a longer period of time and calculated as an average spectrum.

6.2 Spectral Slope

[Slope, Tilt]

Spectral slope refers to the global energy distribution in the power spectrum. Spectra of voiced segments have more energy in the lower frequency bands than in the higher frequency bands; with modal voice, the energy drop in the source spectrum is typically around 12 dB/octave.

This effect is particularly pronounced for breathy voices with an extended open phase, i.e. the drop in energy in the higher frequency bands is more pronounced than with a modal voice and the spectrum is therefore more inclined. With an extended closed phase, on the other hand, there may be an increase in energy in the higher frequency ranges and the spectrum is less inclined.

Several methods for measuring spectral tilt are described in the literature.⁵ **VOXplot** uses two of them: (1) the global energy distribution in the Long-Term Average Spectrum (*Slope*) and (2) the slope of the regression line through an LTAS (*Tilt*). Both parameters are also included in the *Acoustic Voice Quality Index (AVQI)*, which is presented in chapter 7.1.

Slope is calculated as the energy difference between two arbitrary frequency bands. To be compatible with the AVQI algorithm, **VOXplot** uses 0 Hz to 1000 Hz as the low frequency band, and 1 kHz to 10 kHz as the high frequency band. If the energy decreases towards the higher frequencies, a negative *Slope* value is obtained, i.e. for human voices a *slope* value < 0 is always obtained. More strongly inclined spectra result in smaller values (e.g. -30 dB) than less inclined spectra (e.g. -10 dB). As a very rough (!) guide, a *Slope* value between -10 and -20 dB can be expected for modal voices; however, values outside of this range are sometimes also found for healthy voices; similarly, values inside of this range are found for pathological voices. On its own, a single *Slope* value is therefore hardly meaningful. However, if you compare two measurements, an increase from -35 dB to -20 dB, for example, can objectively be considered an improvement.

The *Tilt* parameter describes the slope of the regression line through the LTAS. As with the *Slope* parameter, the steeper the slope of the line, the smaller the *Tilt* value (strong slope e.g. -14 , weak slope e.g. -6).

6.3 Amplitude Difference H1-H2

[H1H2]

Voice is produced by more or less regular opening and closing of the glottis. Each phonation cycle consists of an open and a closed phase of the glottis. The temporal proportion of these phases in the overall cycle correlates with voice quality: In modal voices, the open phase is usually slightly longer than the closed phase. A breathy voice has very long open phases and short closed phases, whereas in laryngealization (*creaky voice*) or harsh voices, the closed phase is often elongated, which usually leads to the impression of a rough voice. If an EGG measurement⁶ is available, the phase durations can be measured directly and very accurately in the laryngeal excitation signal (Lx signal); this is shown in Fig. 8.

Without EGG, we must rely on the less reliable measurement of acoustic correlates. The amplitude difference between the first and second harmonics in the spectrum is the most stable acoustic correlate for the relationship between open and closed phase. Figure 9 shows the narrowband spectrum of a vowel /a:/; the first 7 harmonics are labeled and numbered consecutively. H1 corresponds to the fundamental frequency (F_0). For this vowel and this speaker, H4 and H7 are amplified by the first and second formants respectively,

⁵The amplitude difference between H1 and H2 discussed in the next section can also be used as a measure of the spectral slope.

⁶Electroglottography (EGG) is a non-invasive method for measuring glottal activity. The electrical conductivity is measured between two surface electrodes that are fixed to the left and right of the thyroid cartilage. When the glottis is closed, the conductivity is high; when the glottis is open, it is low due to the air gap between the vocal folds.

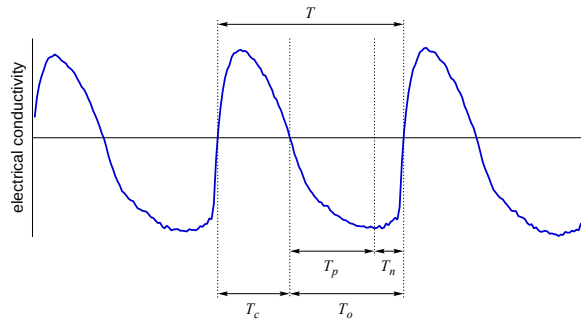


Figure 8: The EGG-generated laryngeal excitation signal (Lx signal). T = period duration of a vocal fold vibration (phonation cycle); T_o = open phase; T_c = closed phase; T_p = increasing airflow; T_n = decreasing airflow.

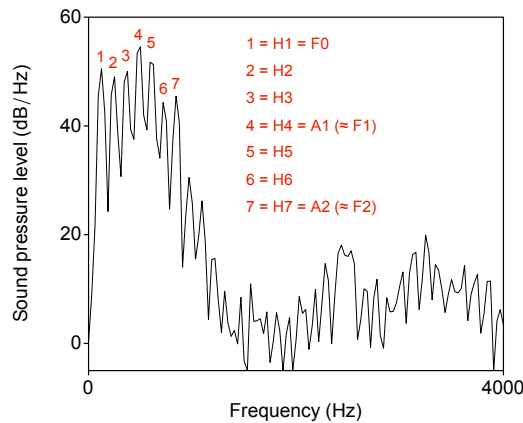


Figure 9: Narrowband spectrum of a vowel /a:/; the first seven harmonics are labeled with the numbers 1-7.

hence the designation A1 and A2.⁷ This amplification effect may affect harmonics other than H4/H7 for other vowels and/or other speakers.

In Fig. 9, H2 has a slightly lower amplitude than H1, indicated by a lower peak. Hence, the amplitude difference H1–H2 gives a small positive value. A longer open phase, represented by a larger open quotient ($\frac{T_o}{T}$), results in a lower H2 amplitude, which in turn leads to a higher positive value for H1–H2 (> 4). Conversely, a smaller open quotient, i.e. an elongated closed phase, leads to an increase in the H2 amplitude. If it exceeds the H1 amplitude, H1–H2 results in a negative value (< -2).

⁷A1 denotes the amplitude of the harmonics in the vicinity of the first formant; A2 denotes the amplitude of the harmonics in the vicinity of the second formant.

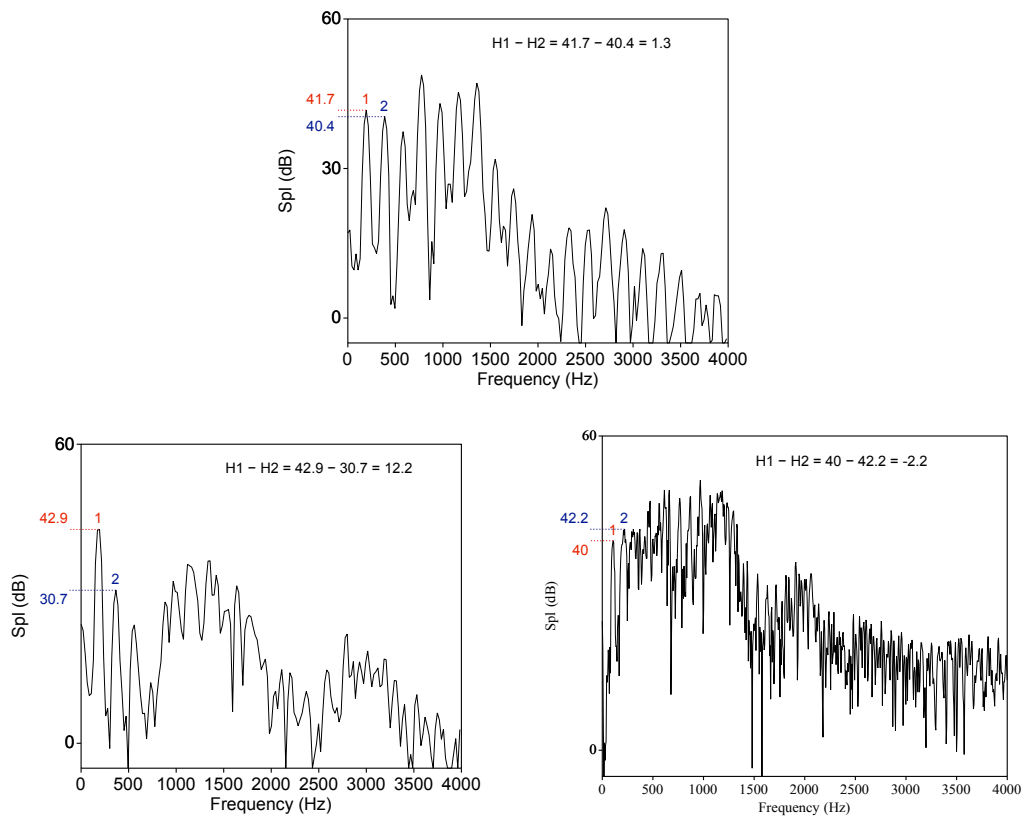


Figure 10: Spectra of 3 different speakers: Normal voice (top), severely breathy voice (bottom left) and moderately harsh voice (bottom right).

The following representation of the H1–H2 value range is only for rough orientation, the thresholds are not properly collected norm data!

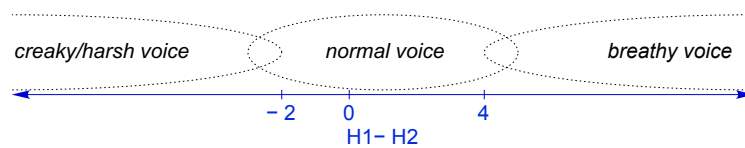


Figure 10 shows an example of a normal voice ($H1-H2 = 1.3$) and two examples of disordered voices: Severely breathy ($H1-H2 = 12.2$) and moderately harsh ($H1-H2 = -2.2$).

Caution: The parameter H1–H2 is very susceptible to interference that has nothing to do with voice quality. The first influencing factor that is relatively easy to avoid is the first vowel formant. For vowels with a low first formant (these are the closed vowels), there is a risk that the first formant raises the amplitude of the low harmonics, which distorts the calculation of the amplitude difference, i.e. the result can no longer be interpreted in terms

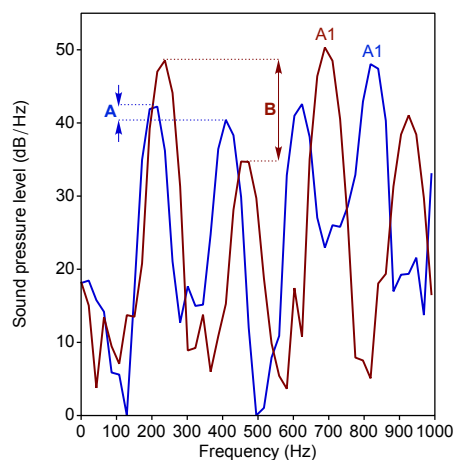


Figure 11: *Spectrum at the beginning (blue) and at the end (red) of the vowel /a:/ in the German word Fahne [fa:nə] (female speaker); $H1-H2 = 2.1$ at the beginning, in the oral part of the vowel (A); in the nasalized part at the end of the vowel, the H1 amplitude is amplified and the H2 amplitude is reduced, $H1-H2 = 13.9$ (B); it can also be observed that the fundamental frequency increases towards the end of the vowel, while the frequency of the first formant decreases during the VC transition.*

of voice quality. Open vowels with a high first formant are therefore best suited as a basis for the measurement; the vowel /a/ is the most suitable.

The second factor that influences the amplitude of the low harmonics is nasality. If a vowel is fully or partially nasalized, the first resonant frequency of the nasal cavity (between 200 and 300 Hz) will usually raise the amplitude of the first (high voices) or second harmonic (low voices). The consequence is the same as above: The result of the amplitude difference between H1 and H2 can no longer be interpreted in terms of voice quality [90]. Even in languages without phonological vowel nasalization, at least partially nasalized vowels are often found as a result of coarticulatory processes in the context of nasal consonants. Furthermore, this measurement obviously cannot be meaningfully applied to dysarthric speakers with hypernasality. Finally, the problem of intrinsic nasality should be mentioned: The more open the vowel, the further lowered the velum [22]. This means that open vowels, which appeared to be an ideal basis for measurement due to their high first formant, have the highest intrinsic nasality of all vowels and must therefore be considered problematic.

Nevertheless, the vowel /a/ is usually used. It is assumed that the tangible advantage of the high first vowel formant outweighs the potential disadvantage of intrinsic nasality. Of course, the measurement must be discarded if a vowel is suspected of nasalization.

To conclude this discussion, Fig. 11 shows the influence of anticipatory nasalization on the relationship between H1 and H2 using the example of the vowel /a:/ before a nasal consonant (from the German word *Fahne* [fa:nə]). The amplitude difference at the beginning, in the non-nasalized part of the vowel, is 2.1 dB, which is typically for a modal voice. At the end of the vowel, when the velum is already lowering, the amplitude of H1 is amplified by the first nasal formant, while the amplitude of H2 is reduced (presumably under the

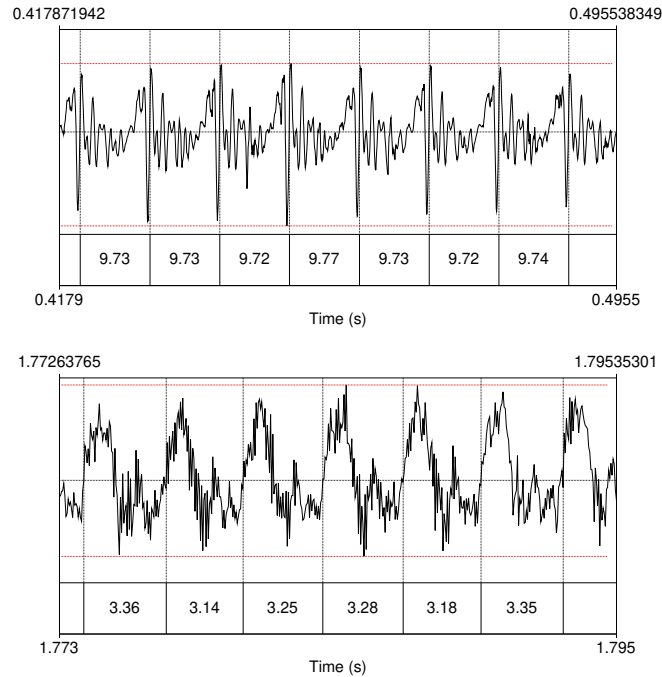


Figure 12: Sections from a sustained vowel: normal voice (top) and hoarse voice (bottom). The period duration in milliseconds is shown below the signal. The maximum amplitude variation within the respective section is marked in red. The maximum deviation in the period duration (jitter) is 0.5% (9.72 ms vs. 9.77 ms) in the upper section compared to 6.6% (3.14 ms vs. 3.36 ms) in the lower section. Deviations in the amplitude range from period to period (shimmer) are visible to the naked eye. (The high-frequency, noisy signal components in the lower section are very pronounced and therefore also easily recognizable; see section 6.5.)

influence of a nasal 'antiformant'). This leads to an amplitude difference of 13.9 dB—in itself a typical value for a severely breathy voice, but in this case solely due to the nasalization.

6.4 Perturbation

[Jitter, Shimmer, PSD]

Phonated speech results in a quasi-periodic acoustic signal. Quasi-periodic means that the signal is periodic enough for a human listener to perceive a harmonic sound and identify a fundamental frequency (i.e. a vocal pitch), but the signal is not periodic in a strict mathematical sense, i.e. the successive periods are not exactly identical. Deviations between neighboring periods (perturbation) affect both the period duration and thus the frequency and the amplitude. The fundamental frequency perturbation is called *Jitter*, the amplitude perturbation *Shimmer* (see Fig. 12). To a small extent, jitter and shimmer are therefore normal phenomena of the human voice. Even slightly increased jitter and shimmer values, e.g. in heavy smokers, are often imperceptible to the ear. Only significantly increased values, as seen in various voice disorders, lead to the impression of a hoarse voice.

The rule of thumb for all jitter and shimmer parameters is that smaller values are better. Therefore, if a patient's voice changes from *Jitter local* = 3.27% at the first examination to *Jitter local* = 0.73% at the follow-up examination, an objective improvement can be

assumed. Healthy voices generally have jitter-local values < 0.5% and shimmer values < 3%.⁸ It should be noted that natural, continuous speech usually results in higher jitter and shimmer values than sustained phonation. The norm values of 0.5% and 3% therefore only apply to sustained phonation.

Jitter can be determined using different algorithms. Very often *Jitter local* is used, the average difference between two neighboring periods divided by the average period. **VOXplot** also calculates the 5-point period perturbation quotient (*Jitter ppq5*), which smoothes over 5 neighboring periods, i.e. works somewhat less locally and can therefore be regarded as a more robust measurement. In [18], *Jitter ppq5* was identified as a valid marker for hoarseness with a cut-off value of 0.29%. Interested parties can find mathematical details in the Praat manual.

Period Standard Deviation (PSD) PSD describes the standard deviation of the period duration in a signal section and thus—similar to jitter—provides information about the frequency perturbation of a quasiperiodic signal. Regular, periodic phonation results in a low standard deviation of the period duration, so the same rule of thumb applies here as above: The smaller the value, the better.

6.5 Harmonicity

[HNR, HNR-D, GNE, HF Noise]

The ratio between harmonic and non-harmonic (noisy) signal components in the quasiperiodic excitation signal is evaluated under the keyword *harmonicity*. A low level of noise can be observed even in a healthy voice, but pronounced noise components are a typical indicator of breathiness or hoarseness in general. **VOXplot** provides three measures for assessing harmonicity: *HNR*, *HNR-D*, and *GNE*.

HNR

The *Harmonics-to-Noise Ratio* (HNR) indicates the ratio of harmonic or periodic components to noise components in the excitation signal. An HNR of 20dB means that 99% of the signal energy is attributable to periodic signal components and 1% to noise components. An HNR of 0% means that the signal energy is due in equal parts to periodic and noise components. Interpretation: The higher the value, the better. In healthy voices, values above 20dB are usually achieved, the **VOXplot** cut-off score is 23.34dB (sustained phonation) [18].

In the case of irregular vocal fold vibrations, as observed in many voice disorders, the classic HNR algorithm is quickly overwhelmed, as it is based on correlation analysis in the time domain. In such cases, the following two parameters offer advantages, especially *Glottal-to-Noise-Excitation Ratio*.

⁸Due to different algorithms for period detection, the jitter measurements and consequently the jitter norm values of **VOXplot**/Praat (0.5%) and MDVP (1.04%) are not comparable [2, 24, 70, 80].

HNR-D

In 1987, Dejonckere and Lebacqz published a method to determine the *Harmonics-to-Noise Ratio* in the spectrum [31]. For this purpose, the harmonic structure in the frequency band between 500 Hz and 1500 Hz is analyzed based on the cepstrally determined fundamental frequency. The frequency band analyzed corresponds approximately to the range between the first and second formants of a sustained /a/. Again, the larger the value, the better; healthy voices generally achieve values well above 20dB.

Glottal-to-Noise-Excitation Ratio (GNE)

Noise components in the excitation signal can be attributed to different causes: (1) perturbation noise (see section 6.4) and (2) additive noise. Perturbation noise is caused by irregular vocal fold vibrations and can be quantified using jitter and shimmer measurements. Measures of harmonicity, on the other hand, aim to quantify the amount of additive noise in the signal. Additive noise is caused by turbulence, such as that caused by incompletely closing vocal folds, and therefore correlates with breathiness (as opposed to jitter/shimmer). As mentioned above, the classic HNR analysis is susceptible to irregular vocal fold vibration (i.e. increased jitter or shimmer values) and is therefore not well suited to analytically separating the two sources of noise. The *glottal-to-noise excitation ratio*, on the other hand, proves to be largely independent of perturbation noise and thus a robust measure of additive noise even with pronounced irregularity of a voice. The calculation method is described in [76]. The larger the value, the better; the **VOX**plot cut-off value is 0.89 (sustained phonation) [18].

High Frequency Noise (HF Noise)

HF Noise is a quotient that describes the ratio between the relative acoustic energy below a reference frequency and the relative acoustic energy above the same reference frequency.[30] **VOX**plot uses Long-Term-Average-Spectra (LTAS) with a reference frequency of 6 kHz to calculate the quotient. Hoarse voices with a strong additive noise component have more energy in the frequency band above 6 kHz (→ smaller quotient) than normal voices (→ larger quotient). The higher the value, the better. The **VOX**plot limit is 2.28 dB (sustained phonation).

6.6 Voice Breaks

[Voice Breaks]

Degree of voice breaks is a measure of the proportion of non-phonated sections in the total duration of the analyzed sound. Values greater than 0 can be considered a deviation from the norm. This value is only calculated if the voice sample is a sustained vowel. In continuous speech, 'voice breaks' occur frequently, including unvoiced segments and pauses. Therefore, this parameter is not applicable in this context.

6.7 Cepstral Peak

[CPPS]

Since the importance of *Smoothed Cepstral Peak Prominence* (CPPS) for acoustic voice analysis was postulated by James Hillenbrand and colleagues in the mid-1990s [47, 48], numerous studies have shown that CPPS is on the one hand very robust and on the other hand, of all the parameters discussed so far, the one with the greatest relevance, i.e. the one that correlates best with auditory voice quality judgments, especially in the breathiness dimension. CPPS is also suitable as an objective measure for determining the severity of dysphonia [4, 16, 27, 42, 45, 63, 64, 71, 73, 78, 83, 88, 89].

CPPS can be understood as a measure of 'spectral noise' and is based on the calculation of a so-called cepstrum. This is an inverse Fourier transformation of a logarithmized spectrum, i.e.—to put it simply—a time signal is first subjected to a spectral analysis and the result is then subjected to a further spectral analysis. The inventors of the method derived the term *cepstrum* from the word *spectrum*—the first four letters are reversed [25]. Accordingly, we do not find the *frequency* plotted on the x-axis but the *quefrequency*, the equivalent of the (*harmonics*) is called *rahmonics* and the whole thing is not an *analysis* but an *alanysis*...

The CPPS measurement utilises the following principle: the greater the periodicity of a time signal, the more distinct the harmonic structure of the resulting spectrum and the more prominent the first peak (rahmonic) of the resulting cepstrum. The amplitude of this cepstral peak thus reflects the intensity of the harmonic structure of the amplitude spectrum: a pronounced harmonic structure results in a prominent cepstral peak; a weakly pronounced harmonic structure with strong 'spectral noise' results in a flat cepstrum or a cepstral peak with low amplitude [41].

Cepstral peak prominence is usually measured in a smoothed cepstrum, hence the name: *Smoothed Cepstral Peak Prominence*. Figure 13 shows for comparison the spectrum, the cepstrum and the smoothed cepstrum of a sustained vowel with a modal voice (left) or with a severely hoarse voice (right; condition after chordectomy with recurrent granulation polyp in the anterior commissure). The difference in periodicity between the two voices is evident from the harmonic grid lines in the spectrum, which are aligned with the fundamental frequency. This is reflected in the prominence of the cepstral peak: in a normal voice, the cepstral peak is clearly distinguishable, while in a hoarse voice, it is barely distinguishable from the surrounding quefrequencies.

The prominence of the cepstral peak is measured as the difference between the amplitude of the peak and the amplitude of the corresponding point (same cepstral quefrequency) on the regression line of the cepstrum; this is illustrated in Figure 14. The graphs display the cepstras of sustained vowels with a modal voice (top), with a moderately hoarse voice (middle) and with a severely hoarse voice (bottom). On the left side of the graphs, the prominence of the cepstral peak (*Cepstral Peak Prominence/CPP*) is plotted as the difference between the peak amplitude and the amplitude of the point on the regression line with the same quefrequency (*peak quefrequency*). The prominence of the smoothed cepstral peak can be obtained by carrying out the same measurement in the smoothed cepstrum and its regression line (plotted on the right of the graph). (*Smoothed Cepstral Peak Prominence/CPPS*).

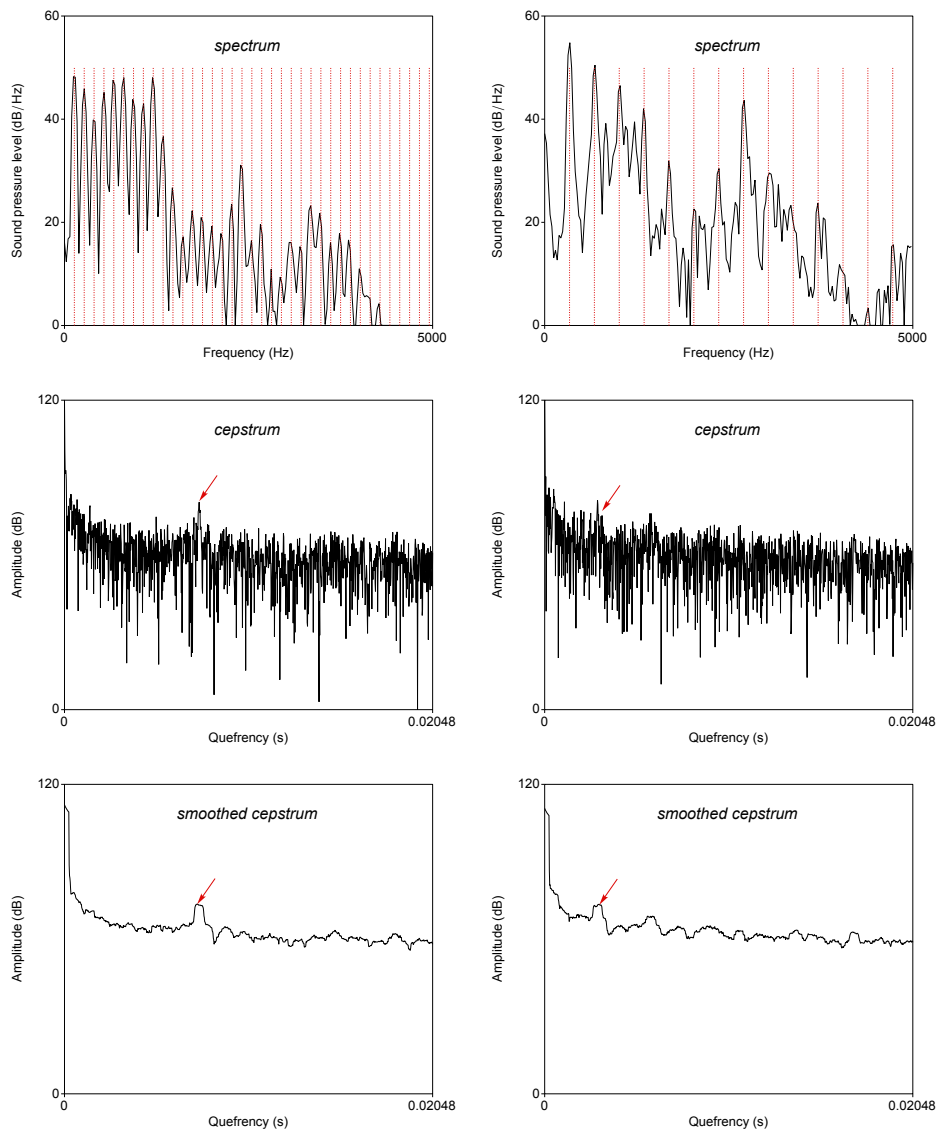


Figure 13: *Spectrum (top), cepstrum (middle) and smoothed cepstrum (bottom) of a sustained vowel; modal voice (left) and hoarse voice (right); the cepstral peak is indicated by red arrows in each case; harmonic grid lines in the spectrum (red) help to assess the periodicity.*

The interpretation of the CPPS value is simple: the higher the value, the better. Healthy voices typically have CPPS values above 14.47 dB [18].

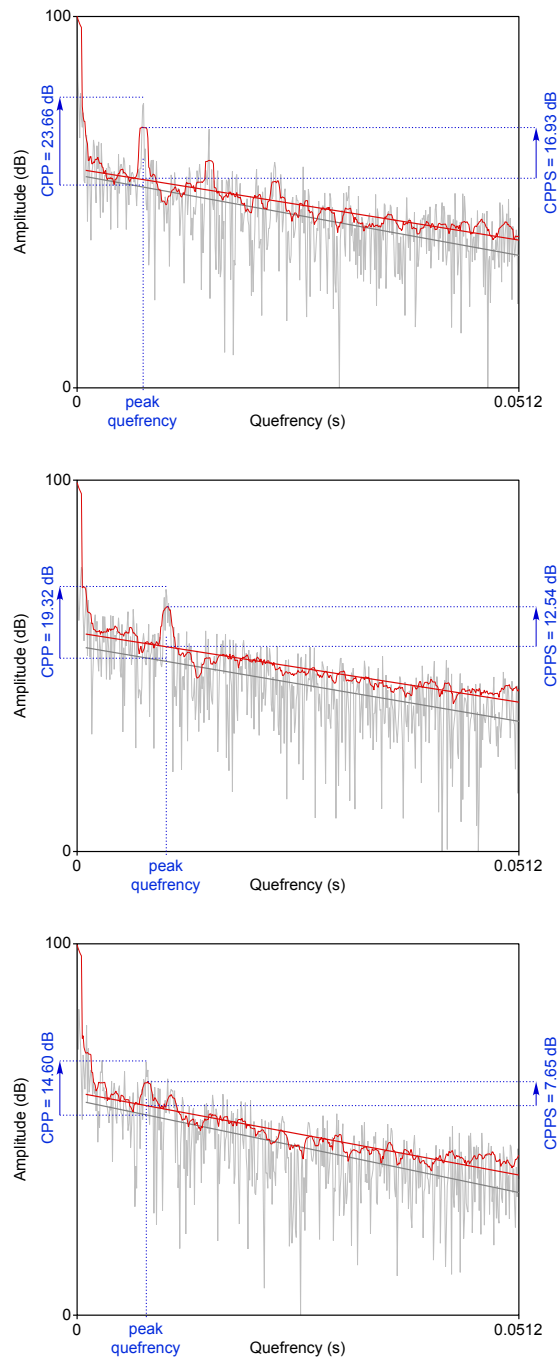


Figure 14: Measurement of (smoothed) cepstral peak prominence (modal voice (top), moderately hoarse (middle), severely hoarse (bottom)); the figure displays the unsmoothed cepstras with regression lines in gray and the smoothed cepstras with regression lines in red.; the Cepstral Peak Prominence (CPP) is plotted on the left, the Smoothed Cepstral Peak Prominence (CPPS) on the right.

6.8 Fundamental Frequency [Pitch mean/min/max/sd, Range]

For cs and sv voice samples (but not for the combined mx voice sample), VOXplot also calculates five simple fundamental frequency (F_0) parameters: mean (*Pitch mean*), minimum and maximum (*Pitch min* and *Pitch max*), the standard deviation of the fundamental frequency (*Pitch sd*) and the pitch range in semitones (*Range*). These parameters only appear in the results output window (see section 4.3.1) and in the CSV export (see section 4.3.3), but not in the graphical voice profile.

Determining the fundamental frequency from the acoustic signal is by no means trivial and there are numerous algorithms with specific advantages and disadvantages dedicated to this task [46]. VOXplot uses a Praat algorithm, which was developed by Paul Boersma [23].

In general, it is important to note that **different fundamental frequency algorithms can produce varying results**. Therefore, algorithmically determined fundamental frequency values should be considered to have a certain degree of uncertainty. This should be taken into account when working with such data. (2) **Fundamental frequency algorithms may produce errors**. Therefore, interpreting algorithmically determined fundamental frequency values requires experience, and the results should always be checked for plausibility.

Some considerations when conducting a plausibility check:

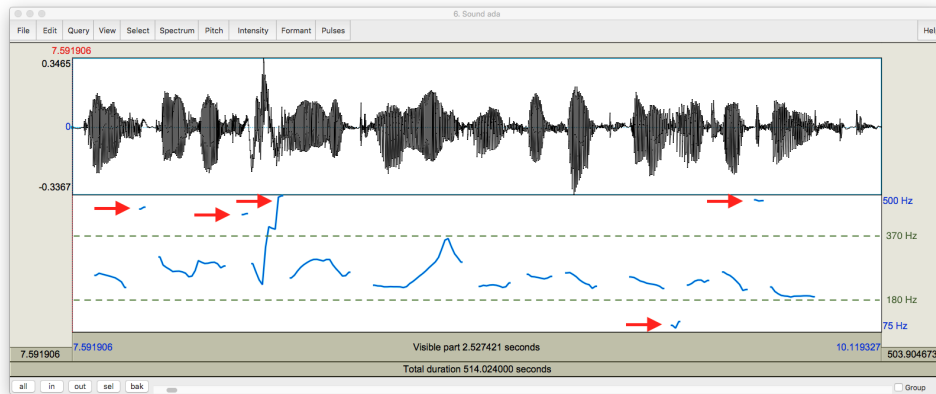
Auditory impression: Does the auditory impression (rising/falling/level pitch) match the algorithmically determined values?

Segmental structure: Fundamental frequency algorithms often struggle with certain segments or segment transitions: (especially voiced) plosives, voiced fricatives, vibrants, sequences of sonorant and vowel at word boundaries (laryngealization), etc.

Phrasal structure: With low ending phrases, many speakers tend to laryngealize; this often leads to unrealistically low F_0 values.

Physiology: Speakers usually have a specific pitch range for speech, which is a section of their entire vocal range determined by physiological conditions. Minor deviations from the typical speaking range are possible but uncommon (and therefore initially ‘suspicious’), while extreme deviations are highly unusual and should be disregarded.

By default, Praat searches for fundamental frequency values in the range between 75 Hz and 500 Hz (analysis range). Although this covers almost all adult speakers, the range is far too wide for most voices, i.e. the algorithm has too many degrees of freedom and is virtually invited to find incorrect extreme values. An example:



The two green dashed lines mark the actual pitch range, which is quite clearly recognizable visually. The red arrows indicate erroneous extreme values. Without further intervention, the Praat algorithm provides unrealistic values in this example, at least for the minimum, maximum and pitch range; the mean value is generally unproblematic. To avoid this, voxplot implements an algorithm proposed by Daniel Hirst [49], which estimates an analysis range adapted to the individual speaker and thus provides more reliable results.

The F_0 parameters and their interpretation: The **mean value** allows conclusions to be drawn about the average speaking pitch of a speaker. If the primary aim is to determine the average speaking pitch, a longer utterance is certainly more suitable than a sustained vowel.

The **standard deviation** is a measure of how much the fundamental frequency deviates from the average fundamental frequency in the course of the analyzed utterance. For a sustained vowel, a smaller value is certainly better (stable pitch), but if a longer utterance is analyzed (e.g. cs sample), a smaller value is more an indication of limited pitch modulation, i.e. monotonous speech.

The **minimum** and **maximum** can be used to derive the **pitch range** used by the speaker. Here, too, a small range of variation is to be rated positively for a sustained vowel (ability to hold the tone), whereas a large range of variation is to be rated positively for a longer utterance ('lively' voice). VOXplot shows the pitch range not in Hertz but in semitones (st); the conversion is done according to the following formula: $st = \frac{12}{\ln(2)} \ln\left(\frac{f_{max}}{f_{min}}\right)$

When reading a text aloud, individuals with a healthy voice should be able to reach at least approximately one octave, which is equivalent to 12 semitones.

In some cases, the evaluation of fundamental frequency parameters may differ from the proposed assessment depending on the goal of voice therapy, for example.

7 Multi-Parametric Voice Quality Indices

It is evident from the previous sections that the acoustic analysis of voice quality is not a simple task. There is not *one* parameter that would correlate clearly and unambiguously with auditory voice quality judgments and be suitable as an objective numerical measure. Therefore, an acoustic voice analysis must always consider several parameters that are interrelated. Multi-parametric indices aim to accomplish this by identifying significant acoustic parameters, weighting them in complex statistical analyses, and combining them into an index.

VOXplot currently incorporates two voice quality indices that have been developed in recent years and whose validity has been demonstrated in numerous studies.: The *Acoustic Voice Quality Index (AVQI)*, which correlates with the overall impression of hoarseness of a voice, and the *Acoustic Breathiness Index (ABI)* as a specific measure of the degree of breathiness of a voice. Unfortunately, an index that covers the roughness dimension does not yet exist [10].

7.1 Acoustic Voice Quality Index

[AVQI]

The Acoustic Voice Quality Index (AVQI) is a multi-parametric, acoustically based index for the objective assessment of hoarseness. The AVQI comprises the following six acoustic parameters: Harmonics-to-Noise Ratio, Shimmer (%), Shimmer (dB), general slope of the long-term average spectrum (Slope), slope of the regression line through the Long-Term Average spectrum (Tilt) and Smoothed Cepstral Peak Prominence (CPPS; CPPS is the main contributor in this model) [8, 68].

For the AVQI analysis, three seconds of a sustained vowel [a:] without onset and offset and a speech sample of continuous speech (usually a passage of phonetically balanced text) are used and analyzed in a combined form. Both voice samples are included in the final AVQI result, as the evaluation of the sustained vowel alone does not have sufficient ecological validity [69]. Since its development in 2010, the AVQI has undergone two important and validated adaptations [21, 52]. First, the structure of the signal processing analysis has been completely combined in Praat (AVQI v.02), instead of using two software programs as before. Second, the duration of the continuous speech sample was standardized language-specifically to the number of syllables, which results in approximately three seconds of voiced segments (AVQI v.03). The two types of voice samples are thus given equal weight to ensure a balanced assessment for the final AVQI result. VOXplot implements the current AVQI v.03.01 which, in addition to the advantages mentioned above, is also significantly improved in terms of diagnostic validity compared to earlier AVQI versions [52].

The AVQI analysis typically yields a value ranging from 0 to 10, with higher values indicating greater hoarseness and vice versa (in rare cases, the limit values of 0 and 10 may be exceeded). The AVQI threshold, i.e. the value that distinguishes between normal and hoarse voices, varies depending on the AVQI version [21, 92] and depending on the language or cultural background [21, 37]. Language-specific validation studies for the AVQI v.03 are

currently available for 12 languages [1, 9, 15, 29, 32, 39, 40, 51, 55, 57, 58, 84, 94]. In the **VOXplot** settings, you can choose between these 12 languages for analysis (see section 5.3).

AVQI has a high clinical utility and meta-analyses show good results in terms of diagnostic accuracy, criterion-related concurrent validity and the ability to detect changes over time after different voice treatments [21, 52]. Diagnostic accuracy ($n = 11696$), criterion-related concurrent validity ($n = 10683$) and the ability to detect changes ($n = 734$) were assessed using varying numbers of voice samples—in one of the largest studies to date in the field of voice quality analysis [21]. The results of this meta-analysis on the validity of the AVQI are the basis for assessing the diagnostic quality of the AVQI, which is described below.

Regarding diagnostic accuracy, the summary receiver operating characteristic curve (ROC) demonstrated an excellent area under the ROC curve (AUC) of 0.937. Sufficiently high values of the pooled sensitivity of 83% (95% confidence interval: 82% - 83%) and the pooled specificity of 89% (95% CI: 88% - 90%) were determined, as well as a positive likelihood ratio of 7.75 (95% CI: 6.04 - 9.95) and a negative likelihood ratio of 0.20 (95% CI: 0.17 - 0.23). In addition, high correlations were determined for the AVQI with a weighted average correlation coefficient of $r = 0.796$ for the validity of the sensitivity to change.

It is not possible to derive a general cut-off score from the data of this meta-analysis, as the pooled sensitivity and specificity from all versions and development steps of the AVQI have been summarized. Therefore, **VOXplot** uses the language-specific cut-off scores determined in the respective validation studies. The cut-off score determined for English with AVQI v.03, for example, is 1.17 [29], i.e. an AVQI value above 1.17 indicates a hoarse voice, whereby the AVQI value increases with the degree of hoarseness.

Other aspects examined in AVQI studies:

- ▷ Gender has no influence on the AVQI value, age may have a minimal influence [19, 52, 54].
- ▷ Acceptable values were found for AVQI when assessing the voice in the context of detecting voice disorders (normal versus organic/functional voice disorders) [20] and also a moderate to high correlation with standardized questionnaires for assessing voice impairments due to voice problems [84, 95].
- ▷ AVQI proves to be robust against fluctuations in room acoustics, background noise and microphone quality [26, 95]. Nevertheless, these factors should be controlled (see chapter 8).
- ▷ Although the AVQI was developed to measure hoarseness, research has shown that it captures voice quality anomalies related to breathiness more strongly than those related to roughness [11]. Further research is needed in this area.
- ▷ Since **VOXplot** contains an integrated, validated Praat version and necessary changes—if necessary—are extensively tested before publication, the concerns expressed in [92] regarding Praat updates that may have an impact on the analysis functions of AVQI are unfounded.

7.2 Acoustic Breathiness Index

[ABI]

The Acoustic Breathiness Index (ABI) is a multi-parametric acoustic index that measures the degree of breathiness in a voice. The ABI uses the same statistical procedure for the development of the index and the same speech material for the analysis (3 seconds of a sustained vowel [a:] without vocal onset and offset and a (language-specific) speech sample of continuous speech) as the AVQI v.03. The assessment scheme on a scale of 0-10 (the higher the ABI value, the more breathy the voice) is also comparable. As opposed to the AVQI, the ABI measures the degree of breathiness in a voice. The index is calculated using the following nine acoustic parameters: CPPS, Jitter %, GNE, Hf Noise, HNR-D, H1-H2, Shimmer %, Shimmer dB and PSD [17].

A meta-analysis confirms that the ABI has high clinical utility and good diagnostic accuracy. [13]. The analysis included 3603 voice samples, 467 of which were healthy voices and 3136 of which were voice-disordered. The results of this meta-analysis serve as a basis for the assessment of the diagnostic quality of the ABI, which is described below.

The pooled sensitivity is 84% (95% CI: 83% - 85%), the pooled specificity is 92% (95% CI: 89% - 94%). The ability to discriminate between the absence and presence of breathiness is excellent with an AUC of 94%. As comparable standards were ensured in all included studies of the meta-analysis, a general, weighted ABI cut-off score of 3.40 could be determined from six languages (sensitivity: 86%, 95% CI: 84% - 87%; specificity: 90%, 95% CI: 88% - 92%). However, as with the AVQI, **VOX**plot employs the cut-off scores established in language-specific validation studies for the ABI, rather than using the general score (e.g. English: 2,35). Validation studies are currently available for 11 languages [1, 14, 15, 17, 29, 32, 38, 50, 53, 56, 58].

Regarding concurrent validity (correlation between ABI and the auditory-perceptual judgement of breathiness), the Spearman rank-order correlation coefficient ranged from $r = 0,746$ to $r = 0,890$ [13]. The ability of the ABI to detect treatment-related changes in voice quality is considered sufficient with correlation values of $r = 0.71$ to $r = 0.88$ [14, 50]. Age, gender and degree of roughness have no significant influence on the ABI analysis, neither on the basis of natural [50] nor on the basis of synthetic voices [11].

8 Recording: Tips and Recommendations

The quality and comparability of an acoustic voice analysis are influenced by numerous factors, many of which are related to the recording setting, i.e. the circumstances under which a voice sample that is to be subjected to an acoustic analysis is recorded. This includes hardware aspects (microphone, microphone amplifier, AD converter), but also, for example, room acoustics, background noise, etc. These points are briefly addressed below, combined with some tips and recommendations (cf. [5, 26]).

From a pragmatic point of view, we can distinguish between two quality levels: (1) Sufficient quality to ensure comparability and reliability of one's own measurements and (2) highest quality to achieve comparability with study results. Level (1) naturally requires less effort and is cheaper to achieve than level (2). Regardless of which quality level you want to invest in, the most important requirement for comparability is: **Do not change the recording setting unless absolutely necessary.** After determining your optimal setting, it is important to document it and consistently adhere to it. This allows for reliable comparisons between different measurement times.

8.1 Hardware

Microphone

Construction: Both, normal microphones (on a stand) and headset microphones are suitable. Normal microphones may be more susceptible to ambient noise (recommendation: less than 50 dBA) and it can be more difficult to control the constant distance to the sound source (≈ 30 cm). Headset microphones are less susceptible to ambient noise and the constant distance to the sound source is guaranteed; the disadvantageous proximity effect is generally compensated for by high-quality headset microphones.

Type: Condenser microphones are clearly preferable to dynamic microphones (they are better in terms of impulse response, precision, frequency range, etc.); ideally a condenser microphone with a sensitivity > 60 dB.

Polar pattern: Cardioid, because of the better focus on the sound source, rather than ambient noise.

Frequency range: Ideally the entire range of the human voice is covered: 20 Hz – 20 kHz.

Frequency response: Should be as flat as possible, i.e. maximum 2 dB deviation between 20 Hz and 8 kHz and ideally up to 20 kHz.

Equivalent noise level: (self-noise of the microphone) max. 25 dB(A)

Maximum sound pressure level: 126 dB(A) or more

As an example, I would like to mention two microphones that meet these requirements and offer good value for around \$ 120: The stationary microphone AKG P120 (tested in [26]) and the headset microphone AKG C 544 L (used for instance in [15, 57]). With these two microphones you can therefore achieve quality level (2) (although you will have to invest in additional accessories, see below). Of course, other manufacturers also offer very good microphones in the \$ 100 to \$ 200 range (Audio-Technica, Behringer, Presonus, Røde etc.). With cheaper microphones (level (1)), you have to accept compromises in the frequency range and frequency response and you largely lose comparability with study results and the norm or standard values that may be postulated there. Bottalico *et al.* demonstrate this by using the example of an inexpensive Logitech headset [26]. The same study concludes that jitter and CPPS are the least dependent on the microphone used, followed by shimmer and HNR. The spectral slope parameter and the multiparametric AVQI are the most dependent on the microphone used.

It is generally recommended to avoid using simple external microphones that connect to the computer via a jack plug. These microphones often do not meet the necessary specifications and the analog signal path in the computer may not achieve the desired quality. Ideally, the computer only has to deal with signals that have already been digitized: In principle, digital signals retain their original quality indefinitely. That's why it is important to be cautious when recording and digitizing; after that, there is little risk of technical issues.

Microphone Amplifier / AD Converter

Condenser microphones must be supplied with a DC voltage (*phantom power*) and the analogue microphone signal must be amplified and digitized (analogue-to-digital conversion/AD conversion). USB microphones have all these functions built in and do not require any additional components. They are connected directly to the computer via USB and deliver a fully digitized signal. The recording is controlled directly on the computer with Praat, Audacity or any other recording software and saved as an audio file. USB microphones are available as simple and inexpensive variants such as the Logitech headset mentioned above, but there are also decent large-diaphragm condenser models, which are often marketed under the name *podcast microphone* (from \$ 50). With inexpensive models, savings are made on the transmission quality of the microphone as well as on the quality of the microphone amplifier and the AD converter. Disadvantages of this design: No replacement/upgrade of individual components. Aside from [26], which included the Logitech headset as a deterrent example, I am not aware of any voice study that would have used a USB microphone; this limits comparability. As a general rule, before using a microphone for acoustic voice analysis, the technical data in the operating instructions should be reviewed to ensure that the technical requirements (see page 51) are met as comprehensively as possible.

High-quality microphones, such as the AKG models mentioned above, are initially only microphones; they require additional components to function. In principle, you can purchase all necessary components (phantom power, microphone amplifier, AD converter) individually and in any quality (and at almost any price) and connect them in series. Less complex, cheaper and more common are so-called USB interfaces. These are small boxes that combine the required functions, i.e. supply the microphone connected via XLR plug with power, amplify and digitize the analogue signal and then pass the digitized signal on to the computer via USB. Even the most affordable devices (starting at around \$ 100) from reputable manufacturers such as Audient, Focusrite, Native Instruments, Steinberg etc. deliver excellent quality. Unfortunately, I am not aware of a test of such interfaces with a special focus on acoustic voice analysis, as is the case for microphones ([26]).

If quality level (1) is sufficient, a portable digital recorder (from Zoom, Tascam etc.) may also be a viable option. This allows for high-quality recordings to be made independently of the computer. These recorders are powered by (rechargeable) battery and combine a proper condenser microphone, microphone amplifier, AD converter and storage medium in one device and can therefore be used quite universally. Costs: from approx. \$ 80 upwards.

Regardless of the hardware and software used, you should follow these recommendations:

- ▷ sampling rate: 44.1kHz are common
- ▷ bit depth: at least 16 bit
- ▷ file format: an uncompressed PCM format, e.g. the wav format (avoid using formats that compress with lossy compression, such as MP3)

8.2 Room Acoustics

Reduce Ambient Noise

Noise of any kind must be avoided during recording. Sources of ambient noise include computer fans, power supply units (e.g. from desk lamps/screens), paper, footfall sounds, blinds, fax machines, etc. Such sources of noise should either be avoided altogether (e.g. placing reading texts on a music stand) or at least shielded (polystyrene sheets, cushions, etc.) or their influence should be minimized by keeping as much distance as possible from the microphone. Recommendation: The difference between the ambient noise level and the level of the voice signal (signal-to-noise ratio) should be at least 30 dB(A), preferably more. The signal-to-noise ratio of a specific recording setting can be easily evaluated using VOXplot; more information on this can be found in section 4.2.1.

Reduce Reverberation

To avoid reverberation, it is recommended to not place speakers directly in front of a wall or screen. Additionally, it is preferable to choose a recording room with carpeting and as few glass surfaces as possible, while also incorporating 'soft' furniture with high sound absorption.

8.3 Speaking Behavior

Last but not least, when recording voice samples, ensure that the volume of the sustained vowel phonation is comparable to the person's habitual sound pressure level. The habitual sound pressure level is best determined during continuous speech (e.g. *The North Wind and the Sun*). If the sustained vowel is phonated significantly louder, the result of the acoustic (and auditory) analysis may alter drastically [60, 98].

8.4 Face Masks

A meta-analysis based on 9 studies (with a total of 422 subjects with and without voice disorders) concludes that wearing a respiratory protective mask during an acoustic voice analysis does not significantly affect the measurement results [12]. The meta-analysis considered 6 acoustic parameters (jitter, shimmer, HNR, CPPS, AVQI and habitual sound pressure level (SPL)) and 2 mask types (medical face mask resp. surgical mask and FFP2 mask).

However, the indirect comparison between the two mask types revealed significant differences in 4 out of 6 parameters (jitter, HNR, CPPS and SPL). Therefore, if wearing a mask is necessary or desirable, recordings can be used without further consideration, but it is important to ensure that the same type of mask is always used.

References

- [1] M. Aghajanzadeh, S. Saeedi, S. Jalaie, K. Esarian, and B. Barsties v. Latoszek. "Validation of the Acoustic Voice Quality Index and the Acoustic Breathiness Index in the Persian Language". In: *Journal of Voice* (2023, in press).
- [2] O. Amir, M. Wolf, and N. Amir. "A Clinical Comparison between Two Acoustic Analysis Softwares: MDVP and Praat". In: *Biomedical Signal Processing and Control* 4.3 (2009), pp. 202–205.
- [3] S. N. Awan, A. Giovinco, and J. Owens. "Effects of Vocal Intensity and Vowel Type on Cepstral Analysis of Voice." In: *Journal of Voice* 26.5 (2012), 670e15–670e20.
- [4] S. N. Awan, N. Roy, and S. M. Cohen. "Exploring the Relationship between Spectral and Cepstral Measures of Voice and the Voice Handicap Index (VHI)." In: *Journal of Voice* 28.4 (2014), pp. 430–439.
- [5] B. Barsties and M. De Bodt. "Assessment of Voice Quality: Current State-of-the-Art". In: *Auris, Nasus, Larynx* 42.3 (2015), pp. 183–188.
- [6] B. Barsties, U. Hoffmann, and Y. Maryn. "Spektrografische Stimmtypenklassifizierung Zur Beurteilung Der Stimmqualität". In: *Laryngo-Rhino-Otol* 95 (2016), pp. 105–111.
- [7] B. Barsties and Y. Maryn. "Der Acoustic Voice Quality Index in Deutsch. Ein Messverfahren Zur Allgemeinen Stimmqualität". In: *HNO* 60.8 (2012), pp. 715–720.
- [8] B. Barsties and Y. Maryn. "External Validation of the Acoustic Voice Quality Index Version 03.01 With Extended Representativity". In: *Annals of Otolaryngology & Laryngology* 125.7 (2016), pp. 571–583.
- [9] B. Barsties and Y. Maryn. "The Improvement of Internal Consistency of the Acoustic Voice Quality Index". In: *American Journal of Otolaryngology* 36.5 (2015), pp. 647–656.
- [10] B. Barsties v. Latoszek, M. De Bodt, E. Gerrits, and Y. Maryn. "The Exploration of an Objective Model for Roughness with Several Acoustic Markers". In: *Journal of Voice* 32.2 (2018), pp. 149–161.
- [11] B. Barsties v. Latoszek, M. Englert, J. C. Lucero, and M. Behlau. "The performance of the Acoustic Voice Quality Index and Acoustic Breathiness Index in synthesized voices". In: *Journal of Voice* (2021, in press).
- [12] B. Barsties v. Latoszek, V. Jansen, C.R. Watts, and S. Hetjens. "The Impact of Protective Face Coverings on Acoustic Markers in Voice: A Systematic Review and Meta-Analysis". In: *Journal of Clinical Medicine* 12.18 (2023).
- [13] B. Barsties v. Latoszek, G.-H. Kim, J. Delgado Hernández, K. Hosokawa, M. Englert, K. Neumann, and S. Hetjens. "The validity of the Acoustic Breathiness Index in the evaluation of breathy voice quality: A Meta-Analysis". In: *Clinical Otolaryngology* 46.1 (2021), pp. 31–40.
- [14] B. Barsties v. Latoszek and B. Lehnert. "Interne Validität des Acoustic Voice Quality Index Version 03.01 und des Acoustic Breathiness Index". In: *Laryngo-Rhino-Otologie* 97.09 (2018), pp. 630–635.

- [15] B. Barsties v. Latoszek, B. Lehnert, and B. Janotte. "Validation of the Acoustic Voice Quality Index Version 03.01 And Acoustic Breathiness Index in German". In: *Journal of Voice* 34.1 (2020), 157.e17–157.e25.
- [16] B. Barsties v. Latoszek, Y. Maryn, E. Gerrits, and M. De Bodt. "A Meta-Analysis: Acoustic Measurement of Roughness and Breathiness". In: *Journal of Speech, Language, and Hearing Research* (2018), pp. 1–26.
- [17] B. Barsties v. Latoszek, Y. Maryn, E. Gerrits, and M. De Bodt. "The Acoustic Breathiness Index (ABI): A Multivariate Acoustic Model for Breathiness". In: *Journal of Voice* 31.4 (2017), 511-e11-511-e27.
- [18] B. Barsties v. Latoszek, J. Mayer, C.R. Watts, and B. Lehnert. "Advances in Clinical Voice Quality Analysis with VOXplot". In: *Journal of Clinical Medicine* 12.4644 (2023).
- [19] B. Barsties v. Latoszek, N. Ulozaitė-Staniene, Y. Maryn, T. Petrauskas, and V. Uloza. "The Influence of Gender and Age on the Acoustic Voice Quality Index and Dysphonia Severity Index: A Normative Study". In: *Journal of Voice* 33.3 (2019), pp. 340–345.
- [20] B. Barsties v. Latoszek, N. Ulozaitė-Staniene, T. Petrauskas, V. Uloza, and Y. Maryn. "Diagnostic Accuracy of Dysphonia Classification of DSI and AVQI". In: *The Laryngoscope* 129 (2019), pp. 692–698.
- [21] C. Batthyany, B. Barsties v. Latoszek, and Y. Maryn. "Meta-analysis on the validity of the Acoustic Voice Quality Index". In: *Journal of Voice* (2022, in press).
- [22] F. Bell-Berti. "Understanding velic motor control: studies of segmental context". In: *Nasals, Nasalization, and the Velum*. Ed. by M. K. Huffman and R. A. Krakow. Academic Press, 1993, pp. 63–85.
- [23] P. Boersma. "Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound". In: *Proceedings of the Institute of Phonetic Sciences*. Vol. 17. 1193. Amsterdam. 1993, pp. 97–110.
- [24] P. Boersma. "Should Jitter Be Measured by Peak Picking or by Waveform Matching?" In: *Folia Phoniatrica et Logopaedica* 61.5 (2009), pp. 305–308.
- [25] B. P. Bogert, M. J. R. Healy, and J. W. Tukey. "The Quefrency Alalysis of Time Series for Echoes: Cepstrum, Pseudo-Autocovariance, Cross-Cepstrum, and Saphe Cracking". In: *Proceedings of the Symposium on Time Series Analysis*. Ed. by M. Rosenblatt. Wiley, 1963, pp. 209–243.
- [26] P. Bottalico, J. Codino, L. C. Cantor-Cutiva, K. Marks, C. J. Nudelman, J. Skeffington, R. Shrivastav, M. C. Jackson-Menaldi, E. J. Hunter, and A. D. Rubin. "Reproducibility of Voice Parameters: The Effect of Room Acoustics and Microphones". In: *Journal of Voice* 34.3 (2020), pp. 320–334.
- [27] L. F. Brinca, A. P. F. Batista, A. I. Tavares, I. C. Gonçalves, and M. L. Moreno. "Use of Cepstral Analyses for Differentiating Normal from Dysphonic Voices: A Comparative Study of Connected Speech versus Sustained Vowel in European Portuguese Female Speakers." In: *Journal of Voice* 28.3 (2014), pp. 282–286.

- [28] M. Brockmann, C. Storck, P. N. Carding, and M. J. Drinnan. "Voice Loudness and Gender Effects on Jitter and Shimmer in Healthy Adults". In: *Journal of Speech, Language, and Hearing Research* 51 (2008), pp. 1152–1160.
- [29] A. Castillo-Allendes, J. Codino, L.C. Cantor-Cutiva, C.J. Nudelman, A.D. Rubin, B. Barsties v. Latoszek, and E.J. Hunter. "Clinical Utility and Validation of the Acoustic Voice Quality and Acoustic Breathiness Indexes for Voice Disorder Assessment in English Speakers". In: *Journal of Clinical Medicine* 12.7679 (2023).
- [30] P. Dejonckere. "Recognition of hoarseness by means of L. T. A. S." In: *International Journal of Rehabilitation Research* 6.3 (1983), pp. 343–344.
- [31] P. Dejonckere and J. Lebacqz. "Harmonic emergence in formant zone of a sustained [a] as a parameter for evaluating hoarseness". In: *Acta oto-rhino-laryngologica Belgica* 41.6 (1987), pp. 988–996.
- [32] J. Delgado Hernandez, N.M. Leon Gomez, A. Jiménez, L.M. Izquierdo, and B. Barsties v. Latoszek. "Validation of the Acoustic Voice Quality Index Version 03.01 and the Acoustic Breathiness Index in the Spanish Language". In: *Annals of Otology, Rhinology & Laryngology* 127.5 (2018), pp. 317–326.
- [33] D. D. Deliyski, M Evans, and H Shaw. "Influence of Data Acquisition Environment on Accuracy of Acoustic Voice Quality Measurements". In: *Journal of Voice* 19.2 (2005), pp. 176–186.
- [34] D. D. Deliyski, H. S. Shaw, and M. K. Evans. "Adverse Effects of Environmental Noise on Acoustic Voice Quality Measurements". In: *Journal of Voice* 19.1 (2005), pp. 15–28.
- [35] D. D. Deliyski, H. S. Shaw, and M. K. Evans. "Influence of Sampling Rate on Accuracy and Reliability of Acoustic Voice Analysis". In: *Logopedics Phoniatrics Vocology* 30.2 (2005), pp. 55–62.
- [36] D. D. Deliyski, H. S. Shaw, M. K. Evans, and Vesselinov R. "Regression Tree Approach to Studying Factors Influencing Acoustic Voice Analysis". In: *Folia Phoniatria Logop* 58 (2006), pp. 274–288.
- [37] M. Englert, B. Barsties v. Latoszek, and M Behlau. "The impact of languages and cultural backgrounds on voice quality analyses". In: *Folia Phoniatria et Logopaedica* 74.2 (2022), pp. 141–152.
- [38] M. Englert, B. Barsties v. Latoszek, Y. Maryn, and M. Behlau. "Validation of the Acoustic Breathiness Index to the Brazilian Portuguese Language". In: *Logopedics Phoniatrics Vocology* 0.0 (2021), pp. 1–7.
- [39] M. Englert, B. Barsties v. Latoszek, Y. Maryn, and M. Behlau. "Validation of the Acoustic Voice Quality Index, Version 03.01, to the Brazilian Portuguese Language". In: *Journal of Voice* 35.1 (2021), 160.e15–160.e21.
- [40] M. Fantini, A. R. Maccarini, A. Firino, M. Gallia, V. Carlino, C. Gorris, M. S. Bisetti, E. Crosetti, and G. Succo. "Validation of the Acoustic Voice Quality Index (AVQI) version 03.01 in Italian". In: *Journal of Voice* 37.4 (2023), 631.e1–631.e6.
- [41] R. Fraile and J. I. Godino-Llorente. "Cepstral Peak Prominence: A Comprehensive Analysis". In: *Biomedical Signal Processing and Control* 14 (2014), pp. 42–54.

- [42] R. Garrett. "Cepstral- and Spectral-Based Acoustic Measures of Normal Voices". MSc- Thesis, University of Wisconsin-Milwaukee, 2013.
- [43] T. Haderlein, C. Moers, B. Möbius, and E. Nöth. "Automatic Rating of Hoarseness by Text-Based Cepstral and Prosodic Evaluation". In: (2012), pp. 573–580.
- [44] B. Halberstam. "Acoustic and Perceptual Parameters Relating to Connected Speech Are More Reliable Measures of Hoarseness than Parameters Relating to Sustained Vowels". In: *ORL* 66.2 (2004), pp. 70–73.
- [45] Y. D. Heman-Ackah, R. T. Sataloff, G. Laureyns, D. Lurie, D. D. Michael, R. Heuer, A. Rubin, R. Eller, S. Chandran, M. Abaza, K. Lyons, V. Divi, J. Lott, J. Johnson, and J. Hillenbrand. "Quantifying the Cepstral Peak Prominence, a Measure of Dysphonia." In: *Journal of Voice* 28.6 (2014), pp. 783–788.
- [46] W. Hess. *Pitch Determination of Speech Signals: Algorithms and Devices*. Vol. 3. Springer, 2012.
- [47] J. Hillenbrand, R. A. Cleveland, and R. L. Erickson. "Acoustic Correlates of Breathy Vocal Quality". In: *Journal of Speech, Language, and Hearing Research* 37.4 (1994), pp. 769–778.
- [48] J. Hillenbrand and R. A. Houde. "Acoustic Correlates of Breathy Vocal Quality: Dysphonic Voices and Continuous Speech". In: *Journal of Speech, Language, and Hearing Research* 39.2 (1996), pp. 311–321.
- [49] D. Hirst. "The analysis by synthesis of speech melody: from data to models". In: *Journal of speech Sciences* 1.1 (2011), pp. 55–83.
- [50] K. Hosokawa, B. Barsties v. Latoszek, C. A. Ferrer-Riesgo, T. Iwahashi, M. Iwahashi, S. Iwaki, C. Kato, M. Yoshida, M. Umatani, A. Miyauchi, N. Matsushiro, H. Inohara, M. Ogawa, and Y. Maryn. "Acoustic Breathiness Index for the Japanese-Speaking Population: Validation Study and Exploration of Affecting Factors". In: *Journal of Speech, Language, and Hearing Research* 62.8 (2019), pp. 2617–2631.
- [51] K. Hosokawa, B. Barsties v. Latoszek, T. Iwahashi, M. Iwahashi, S. Iwaki, C. Kato, M. Yoshida, H. Sasai, A. Miyauchi, N. Matsushiro, H. Inohara, M. Ogawa, and Y. Maryn. "The Acoustic Voice Quality Index Version 03.01 For the Japanese-Speaking Population". In: *Journal of Voice* 33.1 (2019), 125e1–125e12.
- [52] T. Jayakumar and J. J. Benoy. "Acoustic Voice Quality Index (AVQI) in the measurement of voice quality: A systematic review and meta-analysis". In: *Journal of Voice* (2022, in press).
- [53] T. Jayakumar and J. J. Benoy. "Validation of Acoustic Breathiness Index (ABI) in the South Indian Population". In: *Journal of Voice* (2022).
- [54] T. Jayakumar, J. J. Benoy, and H. M. Yasin. "Effect of age and gender on Acoustic Voice Quality Index across lifespan: a cross-sectional study in Indian population". In: *Journal of Voice* 36.3 (2022), 436.e1–436.e8.
- [55] T. Jayakumar, R. Rajasudhakar, and J. J. Benoy. "Comparison and Validation of Acoustic Voice Quality Index Version 2 and Version 3 among South Indian Population". In: *Journal of Voice* (2022).

- [56] E. Kankare and A.-M. Laukkanen. "Validation of the Acoustic Breathiness Index in Speakers of Finnish Language". In: *Journal of Clinical Medicine* 12.7607 (2023).
- [57] E. Kankare, L. Rantala, T. Ikävalko, B. Barsties v. Latoszek, and A.-M. Laukkanen. "Akustisen äänenlaatuindeksin (AVQI) version 03.01 validointi suomenkielisille puhujille". In: *Puhe ja kieli* 40.3 (2020), pp. 165–182.
- [58] G.-H. Kim, B. Barsties von Latoszek, and Y.-W. Lee. "Validation of Acoustic Voice Quality Index Version 3.01 and Acoustic Breathiness Index in Korean Population". In: *Journal of Voice* 35.4 (2021), 660.e9–660.e18.
- [59] A.-M. Laukkanen, T. Syrjä, M. Laitala, and T. Leino. "Effects of two-month vocal exercising with and without spectral biofeedback on student actors' speaking voice". In: *Logoped Phoniatr Vocol* 29 (2004), pp. 66–76.
- [60] E. Lebbe, J. Vestraete, J. DeKlerck, and Y. Maryn. "De invloed van intensiteit op akoestische stemkwaliteitsmetingen". In: *Stem-, Spraak-en Taalpathologie* 18.1 (2013).
- [61] J. M. Lee, N. Roy, E. Peterson, and R. M. Merrill. "Comparison of Two Multiparameter Acoustic Indices of Dysphonia Severity: The Acoustic Voice Quality Index and Cepstral Spectral Index of Dysphonia". In: *Journal of Voice* 32.4 (2017), 515e1–515e13.
- [62] K. Leong, M. J. Hawkshaw, D. Dentchev, R. Gupta, D. Lurie, and R. T. Sataloff. "Reliability of Objective Voice Measures of Normal Speaking Voices." In: *Journal of Voice* 27.2 (2013), pp. 170–176.
- [63] S. Y. Lowell, R. H. Colton, R. T. Kelley, and Y. C. Hahn. "Spectral- and Cepstral-Based Measures during Continuous Speech: Capacity to Distinguish Dysphonia and Consistency within a Speaker." In: *Journal of Voice* 25.5 (2011), e223–32.
- [64] S. Y. Lowell, R. H. Colton, R. T. Kelley, and S. A. Mizia. "Predictive Value and Discriminant Capacity of Cepstral- and Spectral-Based Measures during Continuous Speech." In: *Journal of Voice* 27.4 (2013), pp. 393–400.
- [65] J. K MacCallum, Y. Zhang, and J. Jiang. "Vowel Selection and Its Effects on Perturbation and Nonlinear Dynamic Measures". In: *Folia Phoniatrica et Logopaedica* 63.2 (2011), pp. 88–97.
- [66] Y. Maryn. "Acoustic Measurement of Overall Voice Quality in Sustained Vowels and Continuous Speech". Ghent University, 2010.
- [67] Y. Maryn, M. De Bodt, B. Barsties, and N. Roy. "The Value of the Acoustic Voice Quality Index as a Measure of Dysphonia Severity in Subjects Speaking Different Languages". In: *Eur Arch Otorhinolaryngol* 271 (2014), pp. 1609–1619.
- [68] Y. Maryn, M. De Bodt, and N. Roy. "The Acoustic Voice Quality Index: Toward Improved Treatment Outcomes Assessment in Voice Disorders". In: *Journal of Communication Disorders* 43.3 (2010), pp. 161–174.
- [69] Y. Maryn, P. Corthals, P. Van Cauwenberge, N. Roy, and M. De Bodt. "Toward Improved Ecological Validity in the Acoustic Measurement of Overall Voice Quality: Combining Continuous Speech and Sustained Vowels". In: *Journal of Voice* 24.5 (2010), pp. 540–555.

- [70] Y. Maryn, P. Corthals, M. De Bodt, P. Van Cauwenberge, and D. Deliyski. "Perturbation Measures of Voice: A Comparative Study between Multi-Dimensional Voice Program and Praat". In: *Folia Phoniatrica et Logopaedica* 61.4 (2009), pp. 217–226.
- [71] Y. Maryn, C. Dick, C. Vandenbruaene, T. Vauterin, and T. Jacobs. "Spectral, Cepstral, and Multivariate Exploration of Tracheoesophageal Voice Quality in Continuous Speech and Sustained Vowels." In: *The Laryngoscope* 119.12 (2009), pp. 2384–2394.
- [72] Y. Maryn and N. Roy. "Sustained Vowels and Continuous Speech in the Auditory-Perceptual Evaluation of Dysphonia Severity". In: *J Soc Bras Fonoaudiol.* 24.2 (2012), pp. 107–112.
- [73] Y. Maryn, N. Roy, M. De Bodt, P. Van Cauwenberge, and P. Corthals. "Acoustic Measurement of Overall Voice Quality: A Meta-Analysis." In: *Journal of the Acoustical Society of America* 126.5 (2009), pp. 2619–2634.
- [74] Y. Maryn and D. Weenink. "Objective Dysphonia Measures in the Program Praat: Smoothed Cepstral Peak Prominence and Acoustic Voice Quality Index." In: *Journal of Voice* 29.1 (2015), pp. 35–43.
- [75] Y. Maryn and A. Zarowski. "Calibration of Clinical Audio Recording and Analysis Systems for Sound Intensity Measurement". In: *American Journal of Speech-Language Pathology* 0.0 (2015), pp. 1–11.
- [76] D. Michaelis, T. Gramss, and H. W. Strube. "Glottal-to-noise excitation ratio—a new measure for describing pathological voices". In: *Acta Acustica united with Acustica* 83.4 (1997), pp. 700–706.
- [77] W. Minnema and H.-C. Stoll. "Objektive Computergestützte Stimmanalyse Mit Praat". In: *Forum Logopädie* 22.4 (2008), pp. 24–29.
- [78] C. Moers, B. Möbius, F. Rosanowski, E. Nöth, U. Eysholdt, and T. Haderlein. "Vowel- and Text-Based Cepstral Analysis of Chronic Hoarseness." In: *Journal of Voice* 26.4 (2012), pp. 416–424.
- [79] K. M. Murray, R. T. Sataloff, and Y. Heman-Ackah. "Smoothed Cepstrum Calculation Using Stressed Vowels in Connected Speech to Objectively Measure Dysphonia". In: (2015).
- [80] H. Oğuz, M. A. Kiliç, and M. A. Şafak. "Comparison of Results in Two Acoustic Analysis Programs: Praat and MDVP". In: *Turkish Journal of Medical Sciences* 41.5 (2011), pp. 835–841.
- [81] V. Parsa and D. G. Jamieson. "Acoustic Discrimination of Pathological Voice". In: *Journal of Speech, Language, and Hearing Research* 44.2 (2001), pp. 327–339.
- [82] V. Parsa, D. G. Jamieson, and B. R. Pretty. "Effects of Microphone Type on Acoustic Measures of Voice". In: *Journal of Voice* 15.3 (2001), pp. 331–343.
- [83] E. A. Peterson, N. Roy, S. N. Awan, R. M. Merrill, R. Banks, and K. Tanner. "Toward Validation of the Cepstral Spectral Index of Dysphonia (CSID) as an Objective Treatment Outcomes Measure." In: *Journal of Voice* 27.4 (2013), pp. 401–410.

- [84] T. Pommée, Y. Maryn, C. Finck, and D. Morsomme. “The Acoustic Voice Quality Index, Version 03.01, in French and the Voice Handicap Index”. In: *Journal of Voice* 34.4 (2020), 646.e1–646.e10.
- [85] T. Pommée, Y. Maryn, C. Finck, and D. Morsomme. “Validation of the Acoustic Voice Quality Index, Version 03.01, in French”. In: *Journal of Voice* 34.4 (2020), 646.e11–646.e26.
- [86] M. Pützer and W. J. Barry. “Instrumental Dimensioning of Normal and Pathological Phonation Using Acoustic Measurements”. In: *Clinical Linguistics & Phonetics* 22.6 (2008), pp. 407–420.
- [87] V. Reynolds, A. Buckland, J. Bailey, J. Lipscombe, E. Nathan, S. Vijayasekaran, R. Kelly, Y. Maryn, and N. French. “Objective Assessment of Pediatric Voice Disorders with the Acoustic Voice Quality Index.” In: *Journal of Voice* 26.5 (2012), p. 672.
- [88] C. Sauder, M. Bretl, and T. Eadie. “Predicting Voice Disorder Status From Smoothed Measures of Cepstral Peak Prominence Using Praat and Analysis of Dysphonia in Speech and Voice (ADSV)”. In: *Journal of Voice* 31.5 (2017), pp. 557–566.
- [89] H.-J. Shim, H. R. Jang, H. B. Shin, and D.-H. Ko. “Cepstral, Spectral and Time-Based Analysis of Voices of Esophageal Speakers.” In: *Folia phoniatrica et logopaedica* 67.2 (2015), pp. 90–96.
- [90] A. P. Simpson. “The First and Second Harmonics Should Not Be Used to Measure Breathiness in Male and Female Voices”. In: *Journal of Phonetics* 40 (2012), pp. 477–490.
- [91] A. Sprecher, A. Olszewski, J. J. Jiang, and Y. Zhang. “Updating signal typing in voice: addition of type 4 signals”. In: *The Journal of the Acoustical Society of America* 127.6 (2010), pp. 3710–3716.
- [92] L. Stappenbeck, Ben Barsties v. Latoszek, B. Janotte, and B. Lehnert. “Acoustic Voice Quality Index and Acoustic Breathiness Index as two examples for strengths and weaknesses of free software in medicine”. In: *Biomedical Signal Processing and Control* 59 (2020), p. 101938.
- [93] J. G. Švec and S. Granqvist. “Guidelines for Selecting Microphones for Human Voice Production Research”. In: *American Journal of Speech-Language Pathology* 19.4 (2010), pp. 356–368.
- [94] V. Uloza, T. Petrauskas, E. Padervinskis, N. Ulozaite, B. Barsties, and Y. Maryn. “Validation of the Acoustic Voice Quality Index in the Lithuanian Language”. In: *Journal of Voice* 31.2 (2017), 257e1–257e11.
- [95] N. Ulozaite-Staniene, T. Petrauskas, V. Šaferis, and V. Uloza. “Exploring the feasibility of the combination of Acoustic Voice Quality Index and Glottal Function Index for voice pathology screening”. In: *European Archives of Oto-Rhino-Laryngology* 276.6 (2019), pp. 1737–1745.
- [96] A. P. Vogel and P. Maruff. “Comparison of Voice Acquisition Methodologies in Speech Research”. In: *Behavior Research Methods* 40.4 (2008), pp. 982–987.

- [97] A. P. Vogel and A. T. Morgan. "Factors Affecting the Quality of Sound Recording for Speech and Voice Analysis". In: *International Journal of Speech-Language Pathology* 11.6 (2009), pp. 431–437.
- [98] I. Yanushevskaya, C. Gobl, and A. Ni Chasaide. "Voice quality in affect cueing: does loudness matter?" In: *Frontiers in psychology* 4 (2013).