



VOXplot

Akustische Analyse der Stimmqualität

Einfach | Standardisiert | Intuitiv

Handbuch

Jörg Mayer

voxplot.lingphon.com

03 | 2024

Gewährleistungsausschluss

Wir stellen **VOXplot** so zur Verfügung, wie es ist, ohne irgendeine Gewährleistung, weder ausdrücklich noch implizit, einschließlich – aber nicht begrenzt auf – die implizite Gewährleistung der Marktreife oder der Verwendbarkeit für einen bestimmten Zweck. Wir möchten darauf hinweisen, dass **VOXplot** *kein* zertifiziertes Medizinprodukt ist und daher für die Nutzung innerhalb der EU nur eine Empfehlung für die Verwendung zu Forschungszwecken ausgesprochen werden kann. Bei Nutzung außerhalb der EU gelten die entsprechenden Regelungen des jeweiligen Landes.

Lizenz

Die Software **VOXplot** ist quelloffen, kostenlos, darf auf beliebig vielen Computern uneingeschränkt genutzt werden und steht unter der GNU General Public License v3.0.

Quellcode-Archiv: voxpathlot.lingphon.com/de/download/

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

VOXplot nutzt die ebenfalls GPL-lizenzierte Open-Source-Software Praat v6.0.48 für die akustische Analyse. Die ausführbaren Pakete, die von voxpathlot.lingphon.com/de/download/ heruntergeladen werden können, beinhalten ein ausführbares, binäres Praat-Paket. Den Quellcode von Praat finden Interessierte auf GitHub:

Praat-Quellcode: github.com/praat/praat

Dieses **VOXplot**-Handbuch steht unter der Creative-Commons-Lizenz *Namensnennung - Keine kommerzielle Nutzung - Keine Bearbeitung 3.0 Deutschland* (CC BY-NC-ND), d.h. es kann bei Namensnennung des Autors in unveränderter Fassung zu nicht kommerziellen Zwecken beliebig vervielfältigt und verbreitet werden.

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

Unterstützung



Die erste öffentlich verfügbare Version von **VOXplot** ist im Herbst 2021 erschienen und das Programm wird seitdem aktiv weiterentwickelt. Auch für die Zukunft haben wir Ideen und Pläne, wie wir **VOXplot** noch besser machen können.

Wenn Sie **VOXplot** regelmäßig nutzen und unsere bisherige Arbeit honorieren sowie die Weiterentwicklung unterstützen wollen, freuen wir uns über eine Spende. Der QR-Code führt Sie auf eine Paypal-Seite, wo Sie uns eine einmalige oder eine jährlich wiederkehrende Spende zukommen lassen können. **Bei allen Spender:innen bedanken wir uns herzlich!**

Danksagung

Ich danke Ben Barsties v. Latoszek für seine Initiative und Unterstützung sowie für zahlreiche Diskussionen und den unschätzbaren Input bei der Entwicklung von **VOXplot**. Auch für die Beiträge und Korrekturen zu diesem Handbuch habe ich herzlich zu danken.

Inhaltsverzeichnis

1	Einführung	5
2	Installation, erster Start und Deinstallation	8
2.1	Windows	8
2.2	macOS	9
2.3	Linux	9
2.4	Source-Code	9
2.5	Erster Start	9
2.6	Deinstallation	10
2.7	VOXplot aktualisieren	11
3	Das Hauptfenster	12
4	Der VOXplot-Workflow	13
4.1	1. Schritt: Untersuchungsdaten	13
4.2	2. Schritt: Stimmproben	14
4.2.1	SNR Evaluation	14
4.2.2	Stimmprobe öffnen	17
4.2.3	Stimmprobe aufnehmen	18
4.2.4	Stimmproben editieren	20
4.2.5	Stimmproben prüfen und speichern	22
4.2.6	Stimmproben löschen	23
4.3	3. Schritt: Analyse und Ergebnisse	24
4.3.1	Ergebnisausgabe	26
4.3.2	Stimmprofil	27
4.3.3	Ergebnisse speichern	29
5	Einstellungen	31
5.1	Standard Aufnahmedauer	31
5.2	Standard-Untersucher*in	31
5.3	Sprach-Einstellungen	32
5.4	Formate	32
5.5	Echtzeit Signaldarstellung	33
5.6	Aktives Aufnahmegerät	33

6 Akustische Stimmparameter und ihre Interpretation	34
6.1 Vorbemerkungen	34
6.2 Spektrale Neigung (Slope/Tilt)	35
6.3 Amplitudendifferenz H1-H2 (H1H2)	36
6.4 Perturbation (Jitter/Shimmer/PSD)	40
6.5 Harmonizität (HNR/HNR-D/GNE/HF Noise)	42
6.6 Stimmabbrüche (Voice Breaks)	43
6.7 Cepstraler Gipfel (CPPS)	43
6.8 Grundfrequenz (Pitch/Range)	47
7 Multiparametrische Stimmqualitätsindizes	49
7.1 Acoustic Voice Quality Index (AVQI)	49
7.2 Acoustic Breathiness Index (ABI)	51
8 Recording: Tipps und Empfehlungen	53
8.1 Geräte	53
8.2 Raumakustik	55
8.3 Stimmverhalten	56
8.4 Mund-Nasen-Masken	56
Literatur	57

1 Einführung

Die Analyse akustischer Stimmparameter bildet eine objektiv-quantitative Ergänzung zur herkömmlichen, subjektiv-auditiven Stimmdiagnostik. Doch obwohl der instrumentelle Aufwand gering ist (ein Computer ist i.d.R. vorhanden, ein für die Stimmanalysen geeignetes Mikrofon ist eine überschaubare Investition, s. Vorschläge in Kapitel 8), stellt die Integration der akustischen Stimmanalyse in den stimmtherapeutischen Alltag bisher eine Herausforderung dar. Unserer Erfahrung nach sind die größten Hürden für Anwender von entsprechenden Software-Werkzeugen, dass diese nicht intuitiv zu bedienen sind und keine robuste Benutzerführung bieten. Folglich erfordern diese Werkzeuge eine erhebliche Einarbeitungszeit und provozieren im hektischen Praxisalltag Fehlbedienungen, die sich negativ auf das Analyse-Ergebnis und somit auf die Validität der Messparameter auswirken können. Mit **VOXplot** möchten wir diese Hürde aus dem Weg räumen. Wir haben **VOXplot** mit dem Ziel entwickelt, ein Werkzeug für die Analyse akustischer Stimmparameter bereitzustellen, das auf bewährten und zuverlässigen Algorithmen sowie robuster Evidenz basiert und zugleich unkompliziert und intuitiv bedienbar ist.

Bewährt und zuverlässig

VOXplot wird in enger Zusammenarbeit mit dem wissenschaftlichen Berater Prof. Ben Barsties v. Latoszek entwickelt. Die Analyse von insgesamt 14 akustischen Parametern und zwei multidimensionalen Indizes beruht auf bewährten Praat-Algorithmen¹ und folgt mit wissenschaftlich fundierten Voreinstellungen den Empfehlungen zur Analyse akustischer Stimmparameter, die in den Arbeiten zu den Stimmqualitätsindizes AVQI und ABI publiziert wurden (vgl. Abschnitt 7).

Einfach und intuitiv

Die Benutzeroberfläche von **VOXplot** ist einfach und robust gestaltet und kann ohne lange Einarbeitungszeit genutzt werden. Ein intuitiver Workflow führt mit wenigen Klicks zu einer vollständigen akustischen Analyse der Stimmqualität. Außerdem ist **VOXplot** multilingual. Die Benutzeroberfläche ist aktuell in drei Sprachen verfügbar: Deutsch, Englisch, Niederländisch. Valide Analyseparameter stehen derzeit für 12 Sprachen zur Verfügung: Deutsch, Niederländisch, Englisch, Spanisch, Japanisch, Koreanisch, Brasilianisches Portugiesisch, Malayalam, Kannada, Farsi, Finnisch und Italienisch.

Das Ergebnis einer **VOXplot**-Analyse ist ein übersichtliches Stimmprofil mit allen Untersuchungsdaten und Messwerten auf einer einzelnen Seite (s. Abb. 1). Zusätzlich zu den numerischen Messwerten kann das Ergebnis mithilfe eines Diagramms, das Normabwe-

¹Praat ist eine Analyse-Software, die als De-facto-Standard in der akustisch-phonetischen Forschung betrachtet werden kann. Die ausführbaren **VOXplot**-Pakete für Windows, macOS und Linux beinhalten Praat, die Software muss nicht gesondert installiert werden.

chungen in 6 besonders relevanten Dimensionen farblich hervorhebt, einfach und intuitiv beurteilt werden. Für gehaltene Vokale wird zusätzlich ein Schmalbandspektrum dargestellt.

VOXplot berechnet und dokumentiert zum einen akustische Stimmparameter, die auf der Analyse der Periodizität eines Signals basieren und die man in ähnlicher Form z.B. auch in der verbreiteten, kommerziellen Lösung *Multi-Dimensional Voice Program* (MDVP) von KayPentax wiederfindet. Darüberhinaus werden spektrale und cepstrale Parameter (wie z.B. *Smoothed Cepstral Peak Prominence* (CPPS) oder *Glottal-to-Noise-Excitation Ratio* (GNE)) sowie multiparametrische Stimmqualitätsindizes wie AVQI und ABI (vgl. Kapitel 7) berücksichtigt, die i.d.R. eine größere Aussagekraft hinsichtlich Schweregrad und Differentialdiagnose von Stimmstörungen haben [17, 42, 63, 64, 68, 71, 78].

Die Analyse umfasst die folgenden 14 Stimmqualitätsparameter, die in Kapitel 6 erläutert werden; daneben werden auch 5 Pitch-Parameter berechnet (Mittelwert, Minimum, Maximum, Standardabweichung und Stimmumfang/*Range*):

- ▷ Spektrale Neigung (**Slope**)
- ▷ Neigung der Regressionslinie durch das Spektrum (**Tilt**)
- ▷ Amplituden-Differenz zwischen 1. und 2. Harmonischen im Spektrum (**H1H2**)
- ▷ Frequenzperturbation: **Jitter local (%)** und **Jitter ppq5 (%)**
- ▷ Amplitudenperturbation: **Shimmer (%)** und **Shimmer (dB)**
- ▷ Period Standard Deviation (**PSD**)
- ▷ Harmonics-to-Noise Ratio **HNR**)
- ▷ HNR nach Dejonckere & Lebacqz (**HNR-D**)
- ▷ Glottal-to-Noise-Excitation Ratio (**GNE**)
- ▷ Hochfrequente Geräuschanteile (High-Frequency Noise / **HF noise**)
- ▷ Smoothed Cepstral Peak Prominence (**CPPS**)
- ▷ Stimmabbrüche (**Voice breaks**)

Bei geeigneten Stimmproben – gehaltener Vokal plus gelesener Text – können außerdem die multiparametrischen Indizes *Acoustic Voice Quality Index* (**AVQI**) und *Acoustic Breathiness Index* (**ABI**) berechnet werden.

Abbildung 1 auf der nächsten Seite zeigt das Stimmprofil einer Sprecherin mit heißerer Stimme.

VOXplot – Akustisches Stimmqualitätsprofil

VOXplot v2.0.0

Name:	Frau Monika Mustermann	Datum:	05.05.2021
ID:	123-456-789-0	Uhrzeit:	09:25
Geburtsdatum:	1.1.1970	Untersucher*in:	nn
Sprache:	German (de-DE)	Aufnahmedatum CS:	03.05.2021
Anmerkungen:	Untersuchung vor Therapiebeginn	Aufnahmedatum SV:	03.05.2021

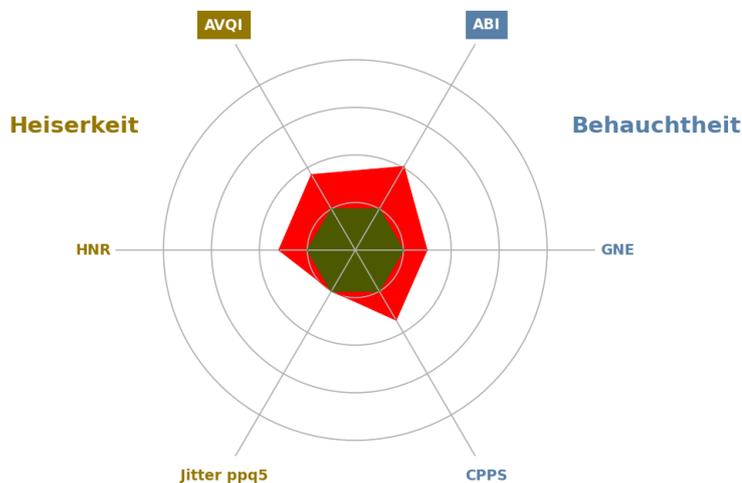
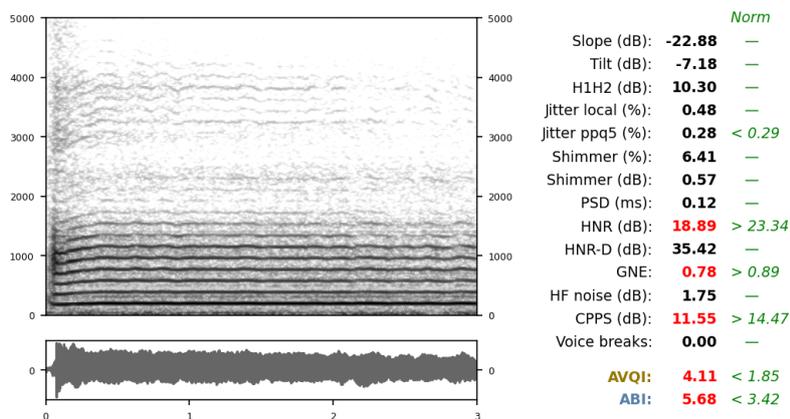


Abbildung 1: Das **VOXplot** Stimmprofil mit 2 Stimmproben: gelesener Text (CS) sowie gehaltener Vokal (SV); Sprecherin mit geringgradiger Heiserkeit. Untersuchungsdaten (oben), Schmalbandspektrogramm der Vokal-Stimmprobe (Mitte links), Messwerte [schwarz/rot] und Normwerte [grün] (Mitte rechts) und Kreisdiagramm mit Normabweichungen (unten).

2 Installation, erster Start und Deinstallation

VOXplot steht als Installationspaket für Windows (32- und 64-bit), als ausführbare Datei für Windows (nur 64-bit) und Linux sowie als macOS-App zur Verfügung. **VOXplot** ist in Python implementiert und quelloffen, d.h. auch der Source-Code kann heruntergeladen und in einer Python-Umgebung ausgeführt werden. Alle Varianten stehen hier zum Download bereit:

voxplot.lingphon.com/de/download/

2.1 Windows

VOXplot ist mit Windows-Versionen ab Version 7 (einschließlich Windows 11) kompatibel.

- ↔ Laden Sie das Installationspaket herunter (`VOXplot64_setup.exe` für 64-bit-Systeme oder `VOXplot32_setup.exe` für 32-bit-Systeme).
- ↔ Nach dem Download starten Sie die Installation mit einem Doppelklick auf das heruntergeladene Paket.
- ↔ Bei einer Neu-Installation wählen Sie zunächst, ob **VOXplot** zentral für alle Benutzer installiert werden soll (Administrator-Rechte erforderlich) oder nur für den/die aktuelle*n Benutzer*in. Bei einer Update-Installation entfällt diese Wahlmöglichkeit, **VOXplot** wird am bestehenden Ort aktualisiert.
- ↔ Als nächstes wählen Sie eine Installationssprache (diese Auswahl gilt nur für den Installationsprozess; die Sprachauswahl für **VOXplot** findet später statt), bestätigen die Lizenz (GPLv3) und entscheiden, ob ein Desktop-Icon angelegt werden soll.
- ↔ Am Ende der Installation können Sie noch entscheiden, ob **VOXplot** sofort gestartet werden soll, dann können Sie das Installationsprogramm beenden und das heruntergeladene Installationspaket löschen.
- ↔ Um **VOXplot** zu starten, doppelklicken Sie auf das Desktop-Icon (sofern vorhanden) oder suchen Sie im Startmenü nach dem **VOXplot**-Eintrag.

Das oben beschriebene Vorgehen ist die bevorzugte Installationsmethode unter Windows. Auf diese Weise wird **VOXplot** gut in das System integriert und startet einigermaßen schnell. Wenn Sie die Systemintegration vermeiden wollen und dafür einen langsameren Programm-Start akzeptieren, dann können Sie auch eine direkt ausführbare **VOXplot**-Datei herunterladen (nur 64-bit). Diese Variante muss nicht installiert werden, sondern wird direkt per Doppelklick gestartet. Durch Löschen der Datei wird **VOXplot** wieder von Ihrem Computer entfernt.

2.2 macOS

Die macOS-App kann mangels Ressourcen bisher nur sehr rudimentär getestet werden und bei der Installation ist mit Problemen zu rechnen (u.a. deshalb, weil Apple der Installation von Software, die nicht aus dem hauseigenen Store stammt, immer mehr Hürden in den Weg legt). Die Lauffähigkeit von **VOXplot** auf aktuellen Geräten mit ARM-Chip (Apple Silicon) können wir mangels Hardware nicht testen – einen Versuch ist es aber wert. Außerdem benötigt die macOS-App *sehr* lange für den Start und ein Splash-Screen, um die Zeit zu überbrücken, wird leider bisher nicht unterstützt.

- ↔ Laden Sie das Archiv `VOXplot.zip` herunter und entpacken Sie es.
- ↔ Die enthaltene App können Sie an einem beliebigen Ort ablegen (z.B. auf dem Desktop).
- ↔ Durch einen Doppelklick auf die App wird **VOXplot** gestartet.
- ↔ Bei Problemen während der Installation beachten Sie bitte die Ttps und Tricks auf unserer Website.

Durch Löschen der App wird **VOXplot** wieder von Ihrem Computer entfernt.

2.3 Linux

VOXplot steht unter Linux ausschließlich für die x86_64-Architektur zur Verfügung.

- ↔ Laden Sie die Datei `VOXplot` herunter und legen Sie sie an einem beliebigen Ort ab.
- ↔ Kennzeichnen Sie die Datei als ausführbar (z.B. im Dateimanager unter *Dateieigenschaften*)
- ↔ Starten Sie **VOXplot** durch Doppelklick auf die heruntergeladene Datei.

Durch Löschen der Datei wird **VOXplot** wieder von Ihrem Computer entfernt.

2.4 Source-Code

- ↔ Laden Sie das Archiv `voxploit_source.zip` herunter und entpacken Sie es.
- ↔ Weitere Hinweise finden Sie in der Datei README, die im zip-Archiv enthalten ist.

2.5 Erster Start

Wenn Sie **VOXplot** zum ersten Mal starten, öffnet sich das **VOXplot**-Einstellungsfenster (Abb. 2). Als Sprache für die Benutzeroberfläche ist *English* voreingestellt. Um auf Deutsch umzustellen, wählen Sie bei der Option **Interface language** *Deutsch* aus (die übrigen Einstellungen

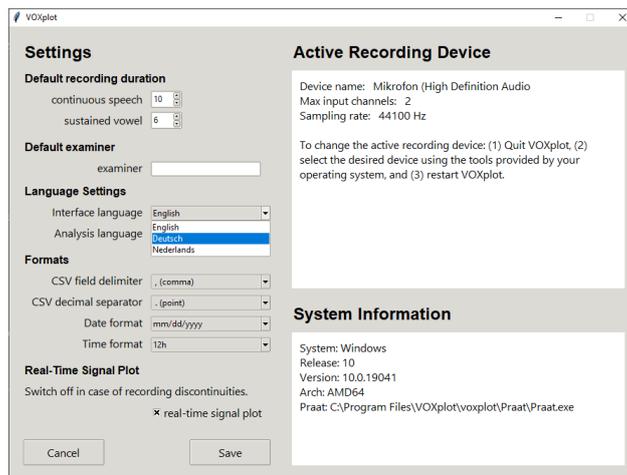


Abbildung 2: Das VOXplot-Einstellungsfenster nach Erst-Installation.

werden in Kapitel 5 erläutert). Klicken Sie dann auf **Save**, beenden Sie VOXplot (**Quit** im Hauptfenster oben rechts) und starten das Programm erneut, erst dann wird die neue Oberflächensprache aktiviert.

VOXplot merkt sich Ihre Einstellungen dauerhaft. Auch bei einem Update und sogar nach einer einfachen Deinstallation bleiben die Einstellungen erhalten, d.h. das englischsprachige Einstellungsfenster mit Standard-Einstellungen bekommen Sie wirklich nur nach dem allerersten Start zu sehen. Wie Sie die Einstellungen im Rahmen einer vollständigen Deinstallation entfernen können, wird im nächsten Absatz beschrieben.

2.6 Deinstallation

Windows

Wenn Sie VOXplot aus einem Installationspaket installiert haben, öffnen Sie die Windows-Einstellungen und dort die App-Übersicht. Suchen Sie in der Liste der installierten Apps nach VOXplot und wählen Sie *Deinstallieren*. Wenn Sie VOXplot nicht installiert haben, sondern eine ausführbare Datei heruntergeladen haben, löschen Sie diese.

Um auch die gespeicherten Einstellungen nach einer Deinstallation oder aus anderem Grund zu löschen, gehen Sie wie folgt vor:

- ↔ Schalten Sie im Explorer unter Ansicht die Anzeige ausgeblendeter Elemente ein
- ↔ Klicken Sie im Schnellzugriff auf Ihren Benutzernamen und navigieren Sie nach `AppData\Local`
- ↔ Löschen Sie das Verzeichnis `voxplot-dir`

macOS & Linux

Löschen Sie einfach die **VOXplot**-App (macOS) bzw. die ausführbare **VOXplot**-Datei (Linux).

Um auch die gespeicherten Einstellungen nach einer Deinstallation oder aus anderem Grund zu löschen, löschen Sie das Verzeichnis `$HOME/.voxplot-dir`

2.7 VOXplot aktualisieren

VOXplot wird aktiv entwickelt und in unregelmäßigen Abständen aktualisiert, um mögliche Fehler zu beseitigen und/oder neue Funktionen zu implementieren. Informationen zu den Änderungen einer neuen Programm-Version gibt es hier:

voxplot.lingphon.com/de/posts/

Ob eine neue Programm-Version zum Download bereit steht, erfahren Sie natürlich auf der Homepage, aber auch innerhalb von **VOXplot** kann geprüft werden, ob ein Update verfügbar ist. Klicken Sie hierzu im Hauptfenster auf . Im Info-Fenster finden Sie den Button . Ein Klick auf diesen Button startet eine Anfrage an unseren Server² und informiert dann, ob Sie bereits die neueste Version nutzen oder ob ein Update verfügbar ist. Falls ein Update verfügbar ist, können Sie die neue Programm-Version von der Download-Seite herunterladen und installieren:

voxplot.lingphon.com/de/download/

²Datenschutz-Information: Bei der Anfrage wird Ihre IP-Adresse an unseren Server übertragen. Dies ist technisch notwendig, um die Anfrage beantworten zu können. Die IP-Adresse wird für max. 7 Tage in den Server-Logs gespeichert und dann gelöscht.

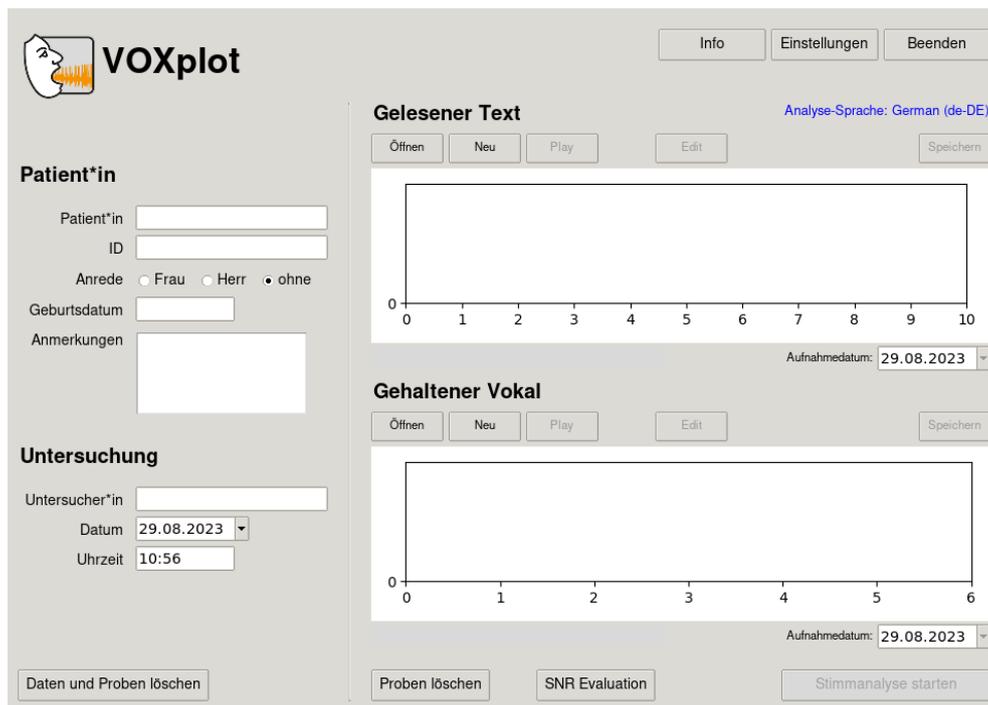


Abbildung 3: Das VOXplot-Hauptfenster.

3 Das Hauptfenster

Nachdem Sie die Benutzeroberfläche auf *Deutsch* umgestellt und VOXplot neu gestartet haben, erscheint das Hauptfenster von VOXplot wie in Abbildung 3.

Das Hauptfenster gliedert sich in drei Bereiche:

- ▷ Oben rechts die drei Buttons **Info**, **Einstellungen** und **Beenden**
- ▷ Links die Eingabefelder für die Untersuchungsdaten
- ▷ Rechts die Steuerungselemente für zwei Stimmproben (Gelesener Text und Gehaltener Vokal)

Der **Info**-Button öffnet ein Fenster mit allgemeinen Informationen zu VOXplot und einem Button, der prüft, ob VOXplot-Updates verfügbar sind (s. Abschnitt 2.7: VOXplot aktualisieren).

Der Button **Einstellungen** öffnet ein Fenster, in dem verschiedene Voreinstellungen vorgenommen werden können und das Sie bereits aus Abschnitt 2.5 kennen (s. Kapitel 5: Einstellungen).

Mit **Beenden** wird VOXplot sofort und ohne weitere Nachfrage beendet.

Die restlichen Bedienelemente lernen Sie im folgenden Kapitel kennen.

4 Der VOXplot-Workflow

Der VOXplot-Workflow besteht aus drei einfachen Schritten, die in den folgenden Abschnitten erläutert werden:

1. Untersuchungsdaten eingeben (optional)
2. Stimmproben bereitstellen (laden oder aufnehmen)
3. Analyse starten

4.1 1. Schritt: Untersuchungsdaten

Im linken Bereich des Hauptfensters können Sie verschiedene Untersuchungsdaten eingeben, wobei sämtliche Angaben optional sind. Die Daten erscheinen – sofern vorhanden – u.a. im Stimmprofil (s. Abb. 1, oben), aber auch in der Ergebnisausgabe (s. Abschnitt 4.3.1) und im CSV-Export (s. Abschnitt 4.3.3).

Die Eingabefelder im Abschnitt **Patient*in** sind für Angaben zur untersuchten Person vorgesehen:

<input type="text" value="Patient*in"/>	:	Name der untersuchten Person
<input type="text" value="ID"/>	:	ID der untersuchten Person (sofern vorhanden)
<input type="text" value="Anrede"/>	:	Anrede der untersuchten Person
<input type="text" value="Geburtsdatum"/>	:	Geburtsdatum der untersuchten Person
<input type="text" value="Anmerkungen"/>	:	Diagnose, Angaben zur Therapie o.ä.

Im nächsten Abschnitt können Angaben zur **Untersuchung** gemacht werden:

<input type="text" value="Untersucher*in"/>	:	Name oder Kürzel der Untersucher*in
<input type="text" value="Datum"/>	:	Datum der Untersuchung
<input type="text" value="Uhrzeit"/>	:	Uhrzeit der Untersuchung

Datum und Uhrzeit werden automatisch auf aktuelle Werte eingestellt, können aber bei Bedarf verändert werden. Wenn in den Einstellungen ein*e Untersucher*in angegeben wurde (s. Abschnitt 5.2), wird auch dieses Feld automatisch ausgefüllt, kann aber ebenfalls verändert werden.

4.2 2. Schritt: Stimmproben

Für eine vollständige Analyse, bei der auch die beiden wichtigen Indizes AVQI und ABI berechnet werden, werden zwei Stimmproben benötigt:

1. Gelesener Text (*continuous speech / cs*)

Der zu lesende Text ist abhängig von der gewählten Analysesprache (s. Abschnitt 5.3). Ist die untersuchte Person deutschsprachig, muss die Analysesprache *German* ausgewählt sein; der Text besteht dann aus den ersten 27 Silben von *Nordwind und Sonne*:

„*Einst stritten sich Nordwind und Sonne, wer von ihnen beiden wohl der Stärkere wäre, als ein Wanderer*“

2. Gehaltener Vokal (*sustained vowel / sv*)

Ein Abschnitt von 3 Sekunden Dauer aus dem gehaltenen Vokal /a/.

Diese Spezifikation der Stimmproben entspricht den Vorgaben zur Berechnung von AVQI und ABI (s. Kapitel 7: Multiparametrische Stimmqualitätsindizes) und nur die Kombination aus gehaltenem Vokal und fortlaufend gesprochenem Text erlaubt eine Stimmbeurteilung "that is truly representative of daily speech and voice use patterns" [69].

Selbstverständlich kann **VOXplot** auch beliebige andere Stimmproben analysieren, **zuverlässige und vergleichbare Ergebnisse werden jedoch nur mit Stimmproben gemäß der obigen Spezifikation erzielt.**

Bereits vorliegende Stimmproben können in **VOXplot** geöffnet werden (Abschnitt 4.2.2). Alternativ können Stimmproben direkt in **VOXplot** aufgenommen werden (Abschnitt 4.2.3). Beide Möglichkeiten können auch kombiniert werden (Stimmprobe 1 wird geöffnet, Stimmprobe 2 wird aufgenommen oder umgekehrt).

Achtung

VOXplot verarbeitet ausschließlich Audio-Daten im WAV-Format, andere Audio-Formate (wie z.B. MP3-Dateien) sind nicht geeignet. Wenn Sie Stimmproben außerhalb von **VOXplot** aufnehmen, achten Sie darauf, diese als WAV-Dateien zu speichern. Stimmproben, die innerhalb von **VOXplot** aufgenommen werden, liegen automatisch im WAV-Format mit geeigneter Abtastrate vor.

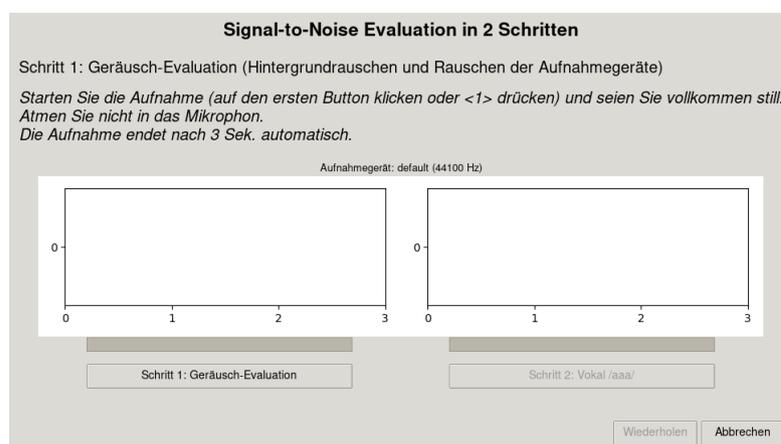
4.2.1 SNR Evaluation

Ein wichtiges Kriterium dafür, ob eine Stimmprobe für die akustische Analyse geeignet ist, ist das Signal-Rausch-Verhältnis (*Signal-to-Noise Ratio*, SNR), das darüber Auskunft gibt, wie gut sich das Nutzsignal (z.B. der gehaltene Vokal) vom Grundrauschen abhebt, das jede Audioaufnahme begleitet. Das Grundrauschen setzt sich zusammen aus dem Hintergrundrauschen (das während der Aufnahme weitgehend reduziert werden sollte) und

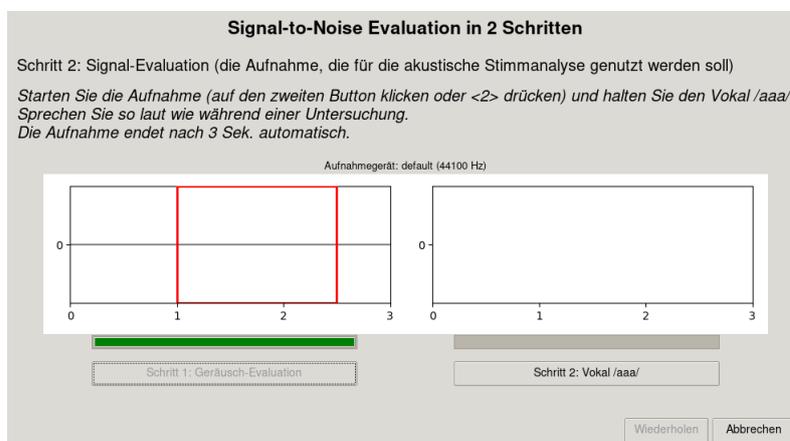
Rauschkomponenten, die von den verwendeten Aufnahmegegeräten (Mikrophon, Verstärker, Wandler usw.) verursacht werden.

VOXplot bietet seit der Version 2.0 die Möglichkeit, SNR für ein konkretes Aufnahme-Setting zu evaluieren. Zum Aufnahme-Setting gehören u.a. der Raum, in dem eine Aufnahme stattfindet, der Abstand zwischen Mikrophon und Sprecher:in sowie die verwendeten Aufnahmegegeräte und deren Einstellungen (z.B. die Aussteuerung des Mikrophons). Um ein Aufnahme-Setting zu evaluieren, richten Sie alles wie für eine routinemäßige Untersuchung ein, starten Sie dann **VOXplot** und klicken Sie den Button **SNR Evaluation** (ganz unten im Hauptfenster). Die SNR-Evaluation besteht aus zwei einfachen Schritten:

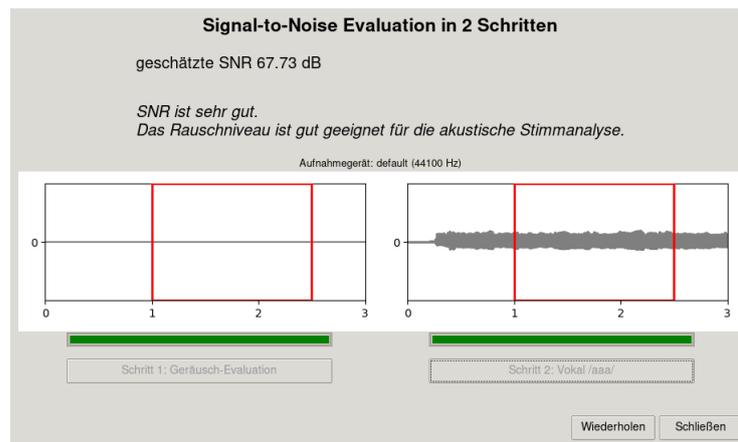
1. Im ersten Schritt soll nur das Grundrauschen aufgenommen werden, d.h. Sie starten eine Aufnahme, indem Sie den Button **Schritt 1: Geräusch-Evaluation** klicken und sind komplett still. Die Aufnahme endet automatisch nach 3 Sekunden und das aufgenommene Signal erscheint.



Für die SNR-Evaluation wird nur der rot markierte Ausschnitt berücksichtigt. Befinden sich unerwünschte Geräusche in diesem Ausschnitt (z.B. ein Ausschlag von einer knallenden Tür o.ä.), sollten Sie die Prozedur abbrechen und wiederholen.



2. Der zweite Schritt besteht aus der Aufnahme eines gehaltenen Vokals unter identischen Bedingungen wie während einer Untersuchung. Starten Sie die Aufnahme indem Sie den Button **Schritt 2: Vokal /aaa/** klicken und produzieren Sie ein /aaa/ (oder lassen Sie eine Patientin oder einen Patienten ein /aaa/ produzieren). Auch hier endet die Aufnahme automatisch nach 3 Sekunden mit der Darstellung des Signals. Oberhalb der Signaldarstellung erscheint das Ergebnis der SNR-Messung und eine Beurteilung der SNR hinsichtlich der Eignung für eine akustische Stimmanalyse (ungeeignet / akzeptabel / sehr gut).



Das Signal-Rausch-Verhältnis wird in dB angegeben. Unbedingt erforderlich für eine valide und reliable akustische Stimmanalyse ist eine $SNR \geq 30$ dB, höchste Akuratheit wird mit einer $SNR > 42$ dB erreicht [5, 34, 75].

Um das Signal-Rausch-Verhältnis zu verbessern, muss das gesamte Aufnahme-Setting geprüft werden:

- ▷ Hintergrundgeräusche reduzieren (Computerlüfter, Gebäudelüftung/Klimaanlagen, fiepsende Netzteile usw.).
- ▷ Wenn kein Headset-Mikrophon verwendet wird: Abstand zwischen Sprecher:in und Mikrophon verringern.
- ▷ Evtl. die Mikrophon-Aussteuerung anpassen.
- ▷ Andere Geräte (Mikrophon, Mikrophonverstärker etc.) verwenden.

Achtung

Die SNR-Evaluation gilt immer nur für ein konkretes Aufnahme-Setting! Spätestens bei einer Änderung des Settings sollte die Evaluation wiederholt werden.

4.2.2 Stimmprobe öffnen

1. Je nachdem, welche Stimmprobe in VOXplot geöffnet werden soll, klicken Sie den Button **Öffnen** im entsprechenden Abschnitt im Hauptfenster (**Gelesener Text** oder **Gehaltener Vokal**).



2. Im Dateiauswahl-Dialog wählen Sie die entsprechende Stimmprobe aus (WAV-Datei).
- 3a. **Gelesener Text** (cs)
Handelt es sich um eine cs-Stimmprobe, wird die Datei in den cs-Editor geladen, um dort ggf. den zu analysierenden Abschnitt (*Einst stritten sich... als ein Wanderer*) aus einer längeren Aufnahme zu extrahieren → **cs-Editor** (S. 20).

- 3b. **Gehaltener Vokal** (sv)

Eine sv-Stimmprobe, die der Vorgabe entspricht (Dauer=3 Sek.) wird direkt in das Hauptfenster übernommen.

Ist die geladene Vokal-Stimmprobe kürzer als 3 Sek., erscheint eine Warnung, dass die sv-Probe nicht für eine standardisierte AVQI- bzw. ABI-Analyse geeignet ist, sie wird aber dennoch geladen und kann analysiert werden. In Abschnitt 4.3 wird erläutert, welche Analysen in diesem Fall durchgeführt werden und welche nicht.

Ist die geladene Stimmprobe länger als 3 Sek., wird sie in den sv-Editor geladen und kann dort sehr einfach durch das Verschieben einer Auswahlbox gekürzt werden → **sv-Editor** (S. 21).

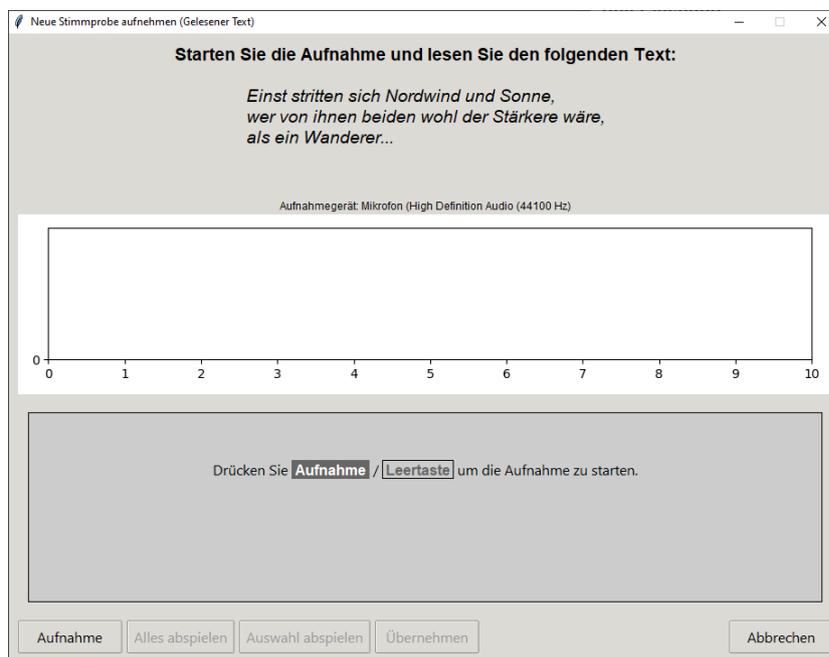
4.2.3 Stimmprobe aufnehmen

1. Je nachdem, welche Stimmprobe aufgenommen werden soll, klicken Sie den Button **Neu** im entsprechenden Abschnitt im Hauptfenster (**Gelesener Text** oder **Gehaltener Vokal**).



2a. Gelesener Text (cs)

Es öffnet sich der cs-Rekorder:



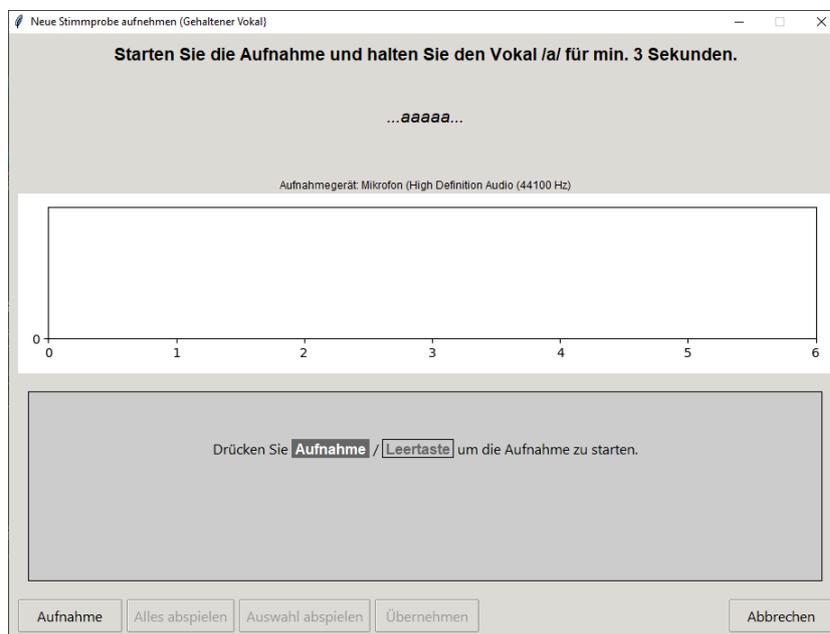
Überprüfen Sie ggf. oberhalb der Signaldarstellung, ob das korrekte Aufnahmegerät ausgewählt ist (s. Abschnitt 5.6). Mit dem Button **Aufnahme** (oder durch Drücken der **Leertaste**) startet die Aufnahme und die untersuchte Person kann den angezeigten Text lesen. Während der Aufnahme wird die Signaldarstellung laufend aktualisiert; dies kann auf weniger leistungsstarken Computern deaktiviert werden (s. Abschnitt 5.5).

Ist die Stimmprobe vollständig, wird die Aufnahme mit dem Button **Stop** (oder erneutes Drücken der **Leertaste**) beendet. Ansonsten endet die Aufnahme automatisch nach Ablauf der voreingestellten Aufnahmedauer (s. Abschnitt 5.1).

Nachdem die Aufnahme beendet ist, wird das Signal im cs-Editor geöffnet und der gewünschte Textabschnitt kann markiert und in das Hauptfenster übernommen werden → **cs-Editor** (S. 20).

2b. Gehaltener Vokal (sv)

Es öffnet sich der sv-Rekorder:



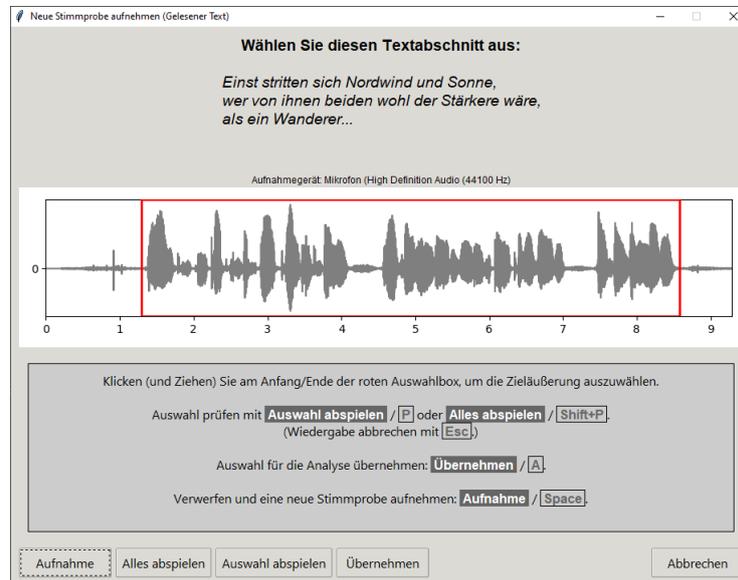
Überprüfen Sie ggf. oberhalb der Signaldarstellung, ob das korrekte Aufnahmegerät ausgewählt ist (s. Abschnitt 5.6). Mit dem Button **Aufnahme** (oder durch Drücken der **Leertaste**) startet die Aufnahme und die untersuchte Person kann aufgefordert werden, den Vokal /a/ möglichst lange zu halten. Während der Aufnahme wird die Signaldarstellung laufend aktualisiert; dies kann auf weniger leistungsstarken Computern deaktiviert werden (s. Abschnitt 5.5).

Ist die Stimmprobe vollständig, wird die Aufnahme mit dem Button **Stop** (oder erneutes Drücken der **Leertaste**) beendet. Ansonsten endet die Aufnahme automatisch nach Ablauf der voreingestellten Aufnahmedauer (s. Abschnitt 5.1).

Nachdem die Aufnahme beendet ist, wird das Signal im sv-Editor geöffnet und der gewünschte Vokalabschnitt kann markiert und in das Hauptfenster übernommen werden → *sv-Editor* (S. 21).

4.2.4 Stimmproben editieren

cs-Editor Der cs-Editor dient dazu, Text-Stimmproben zurecht zu schneiden.



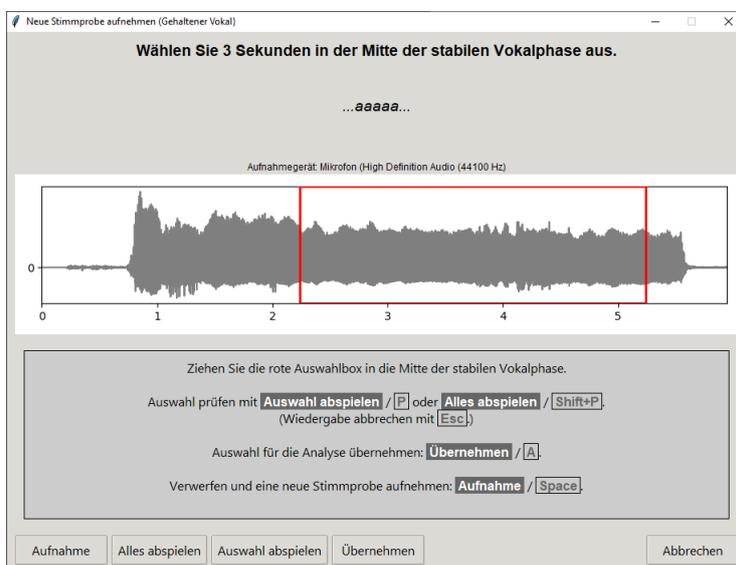
Zunächst ist das Signal, das in den cs-Editor geladen wird, vollständig ausgewählt, d.h. die rote Auswahlbox umfasst das gesamte Signal. Anfang und Ende der der Auswahlbox können nun bei Bedarf mithilfe der Maus (Klicken—Ziehen—Loslassen) so verschoben werden, dass nur der gewünschte Signalabschnitt markiert ist.

Achten Sie darauf, dass Rauschen und Störgeräusche vor bzw. nach dem gewünschten Signalabschnitt außerhalb der Auswahl liegen. Prüfen Sie die Auswahl mit dem Button **Auswahl abspielen**, die gesamte Aufnahme kann mit **Alles abspielen** wiedergegeben werden. Wenn Sie mit der Auswahl einverstanden sind, kann der markierte Signalabschnitt mit **Übernehmen** in das Hauptfenster übernommen und der cs-Editor geschlossen werden.

Die Funktionen der Editor-Buttons können auch per Tastatur gesteuert werden:

Button	Taste(n)
Alles abspielen	Shift + p (Play all)
Auswahl abspielen	p (play)
Wiedergabe vorzeitig abbrechen	Esc
Übernehmen	a (accept)
Abbrechen	Esc
Nur wenn die Aufnahmefunktion aktiv ist:	
Aufnahme	Leertaste

sv-Editor Der sv-Editor dient dazu, Vokal-Stimmproben zurecht zu schneiden.



Durch Verschieben der roten Auswahlbox mithilfe der Maus kann ein geeigneter 3-Sekunden-Abschnitt in einer stabilen Phase des Vokals markiert werden. Die Größe der Auswahlbox ist auf 3 Sekunden fixiert und kann nicht verändert werden, d.h. die Box kann nur insgesamt nach rechts oder links verschoben werden.³

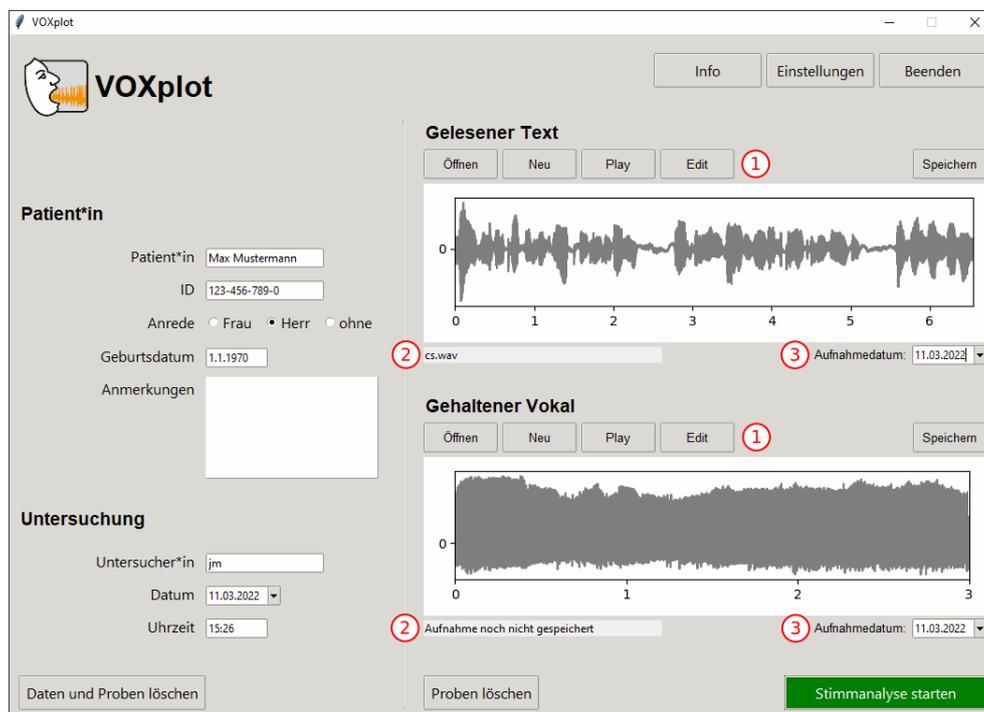
Auch hier stehen die Abspiel-Buttons **Auswahl abspielen** und **Alles abspielen** zur Verfügung, um die Auswahl zu prüfen. Der ausgewählte Signalabschnitt wird mit **Übernehmen** in das Hauptfenster übernommen. Die Tastatursteuerung funktioniert äquivalent zum cs-Editor.

Aufnahme-Button Wird eine Stimmprobe direkt innerhalb von VOXplot aufgenommen, öffnet sich der cs- oder sv-Editor automatisch nach Beenden der Aufnahme. In diesem Fall ist auch der Button **Aufnahme** im jeweiligen Editor aktiv. Sollte sich die Stimmprobe als ungeeignet herausstellen, kann mit diesem Button direkt aus dem Editor heraus eine neue Aufnahme gestartet werden. In allen anderen Fällen ist die Aufnahmefunktion deaktiviert.

Abbrechen-Button Mit **Abbrechen** wird das jeweilige Editor-Fenster geschlossen. Wenn Sie direkt von einer Aufnahme in den Editor gesprungen sind (s. Abschnitt 4.2.3), wird dadurch die aufgenommene Stimmprobe verworfen. Wenn sich der Editor dagegen nach dem Öffnen einer Stimmprobe (s. Abschnitt 4.2.2) oder durch Klicken auf den **Edit** im Hauptfenster (s. Abschnitt 4.2.5) geöffnet hat, wird nur die Auswahl verworfen, die Stimmprobe bleibt erhalten.

³Wenn Sie eine Vokal-Stimmprobe mit anderer Dauer analysieren wollen, müssen Sie diese außerhalb von VOXplot aufnehmen und zurechtschneiden, dann in VOXplot öffnen und den sv-Editor mit **Abbrechen** schließen. Die geöffnete Stimmprobe steht dann unverändert im Hauptfenster zur Analyse bereit.

4.2.5 Stimmproben prüfen und speichern



① Buttons

Wurde eine Stimmprobe geladen oder aufgenommen, werden drei weitere Buttons im Abschnitt der jeweiligen Stimmprobe aktiviert: **Play**, **Edit** und **Speichern**.

Play gibt die jeweilige Stimmprobe wieder, **Edit** öffnet den cs- bzw. sv-Editor und mit **Speichern** kann die jeweilige Stimmprobe dauerhaft auf dem Computer gespeichert werden.

Achtung

Aufnahmen, die direkt in VOXplot erstellt werden, müssen mit dem **Speichern**-Button gespeichert werden, wenn sie dauerhaft erhalten bleiben sollen.

② Dateinamen

Wurde eine Stimmprobe geöffnet, wird unterhalb der Signaldarstellung der Dateiname angezeigt (im Bsp.: *cs.wav* unter der Text-Stimmprobe). Wurde eine Stimmprobe innerhalb von VOXplot aufgenommen und noch nicht gespeichert, wird stattdessen *Aufnahme noch nicht gespeichert* angezeigt (im Bsp.: unter der Vokal-Stimmprobe); sobald die Probe gespeichert wurde, erscheint der Dateiname.

③ **Aufnahmedatum**

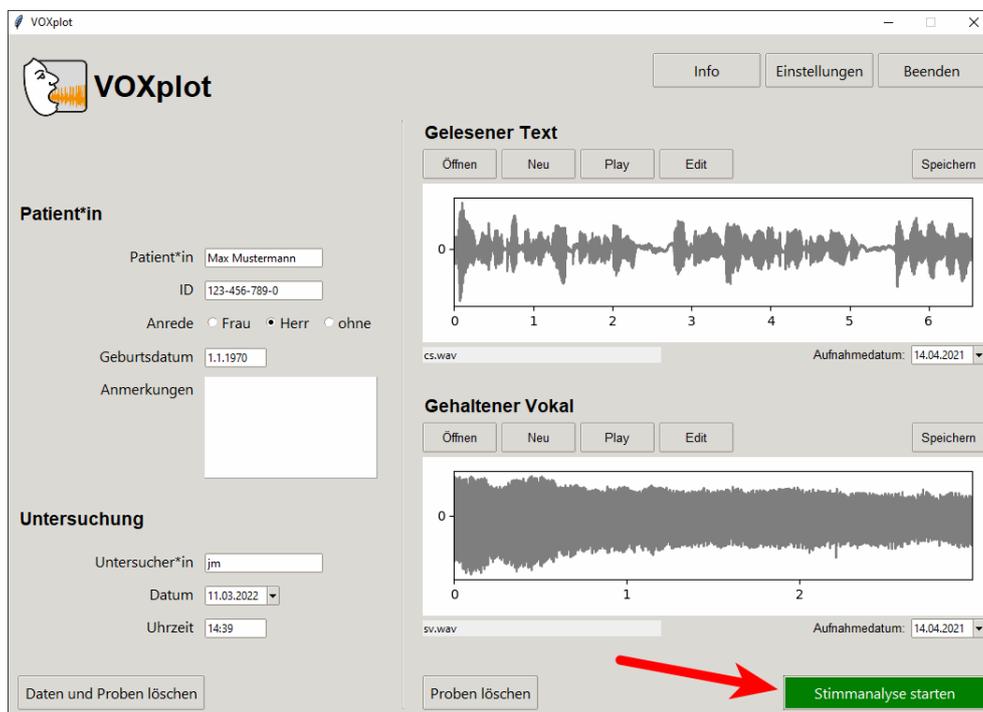
Das Aufnahmedatum jeder Stimmprobe wird automatisch eingetragen. Bei Proben, die bereits auf dem Computer vorlagen und in **VOXplot** geöffnet wurden, wird hier das Erstellungsdatum der Datei übernommen, bei Proben, die innerhalb von **VOXplot** aufgenommen wurden, das aktuelle Datum. Bei Bedarf – z.B. wenn das Erstellungsdatum der Datei nicht mit dem tatsächlichen Aufnahmedatum übereinstimmt – kann das Aufnahmedatum bei beiden Stimmproben verändert werden. Die Aufnahmedaten der beiden Stimmproben erscheinen im Stimmprofil (s. Abb. 1, oben rechts).

4.2.6 Stimmproben löschen

Durch Öffnen (**Öffnen**) oder Aufnehmen (**Neu**) einer Stimmprobe wird eine bereits vorhandene Stimmprobe stets überschrieben. Explizites Löschen der Proben ist somit nicht unbedingt notwendig, aber möglich: Mit dem Button **Proben löschen** werden **beide** Stimmproben aus **VOXplot** entfernt.

Der Button **Daten und Proben löschen** entfernt zusätzlich zu den Stimmproben auch alle Untersuchungsdaten aus dem Formular im linken Bereich und aktualisiert das Datum und die Uhrzeit. Der Button hat somit den gleichen Effekt wie ein Neustart von **VOXplot**.

4.3 3. Schritt: Analyse und Ergebnisse



Sobald mindestens eine Stimmprobe verfügbar ist, wird rechts unten der grüne Button **Stimmenanalyse starten** aktiviert. Wenn sowohl eine Text-Stimmprobe als auch eine Vokal-Stimmprobe mit 3 Sekunden Dauer vorliegt, wird mit diesem Button eine vollständige Analyse gestartet. Liegt nur eine Stimmprobe vor und/oder ist die Vokal-Probe zu kurz oder zu lang, wird nach dem Klicken des Buttons eine Warnung angezeigt. Die Analyse kann dann abgebrochen werden, um die Stimmproben zu korrigieren, oder es kann eine unvollständige Analyse gestartet werden. Die folgende Übersicht zeigt, welche Analysen durchgeführt werden, abhängig von den vorliegenden Stimmproben:

Stimmproben		Analysen			
Text	Vokal (Dauer/Sek.)	num. Parameter	AVQI	ABI	graph. Stimmprofil
✓	✓ (= 3)	✓	✓	✓	✓
✓	✓ (≠ 3)	✓	-	-	-
✓	-	✓	-	-	-
-	✓ (= 3)	✓	-	-	✓
-	✓ (≠ 3)	✓	-	-	-

Achtung

VOXplot prüft nur die Vokaldauer. Ob die Text-Stimmprobe den korrekten Textabschnitt abbildet, kann nicht überprüft werden – es obliegt der Untersucherin bzw. dem Untersucher, darauf zu achten, dass tatsächlich die Textpassage, die von der Einstellung der Analysesprache vorgegeben wird, in die Analyse eingeht.

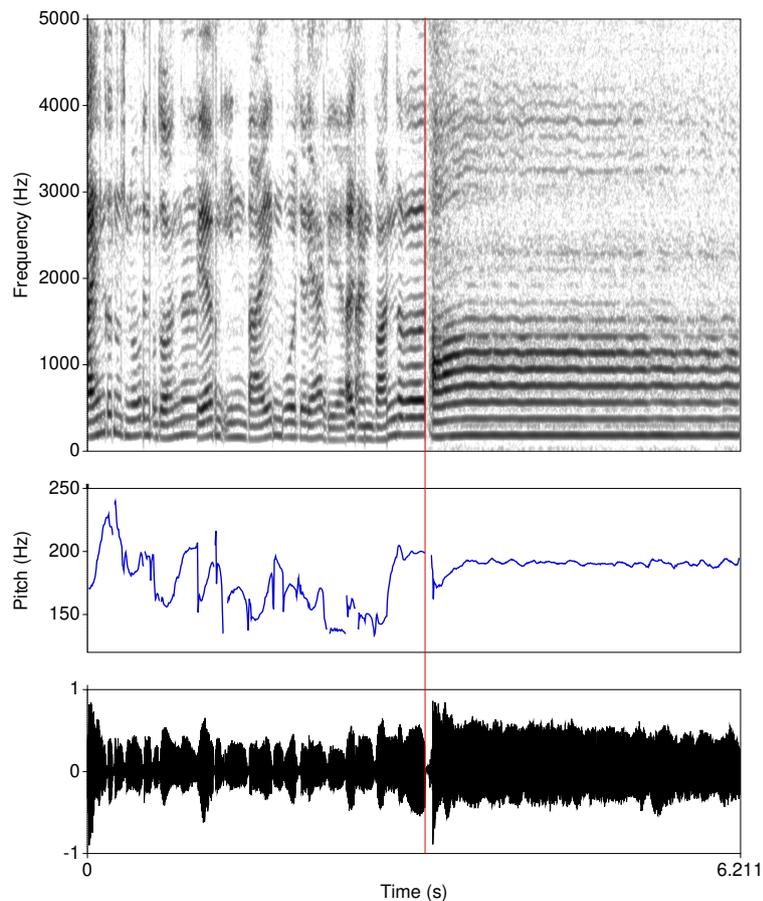


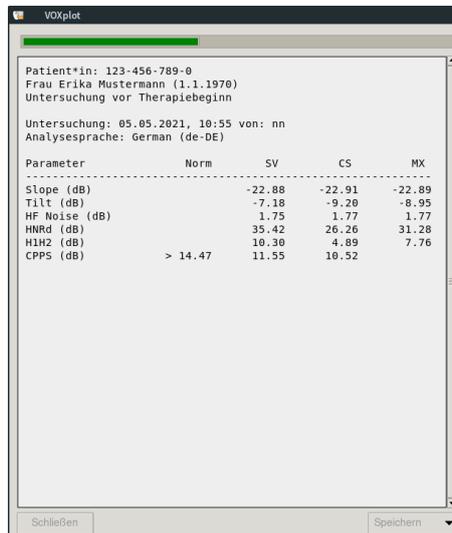
Abbildung 4: *mx*-Stimmprobe; Oszillogramm (unten), Grundfrequenzverlauf (mitte) und Schmalbandspektrogramm (oben). Die senkrechte rote Linie markiert die Grenze zwischen den phonierten Anteilen der *cs*-Stimmprobe (links) und der *sv*-Stimmprobe (rechts). Die Dauer der auf phonierte Anteile reduzierten *cs*-Stimmprobe variiert naturgemäß; im gezeigten Beispiel beträgt sie 3,211 Sek.; zusammen mit der *sv*-Stimmprobe (Dauer: 3 Sek.) ergibt sich somit eine Gesamtdauer der *mx*-Stimmprobe von 6,211 Sek.

Bei einer vollständigen Analyse erzeugt **VOX**plot auf der Basis der beiden vorhandenen Stimmproben (*cs* und *sv*) automatisch eine dritte Stimmprobe (*mx*). Diese besteht aus der vollständigen *sv*-Stimmprobe und aus den phonierten Anteilen der *cs*-Stimmprobe. Der Algorithmus zur Erkennung und Extraktion der phonierten Anteile aus der *cs*-Stimmprobe ist in [81] bzw. [69] beschrieben.⁴ Das Resultat – die *mx*-Stimmprobe – ist im Idealfall eine vollständig phonierte Kombination aus *cs*- und *sv*-Stimmprobe, die insbesondere für die AVQI- und ABI-Analyse verwendet wird (vgl. Abb. 4).

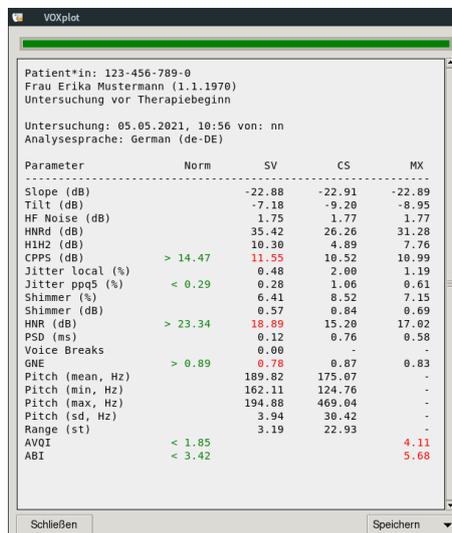
⁴**VOX**plot übernimmt zu großen Teilen die Praat-Implementierung des Algorithmus von Paul Corthals (Universität Ghent, Belgien).

4.3.1 Ergebnisausgabe

Nach dem Start der Analyse erscheint ein Fenster zur Ergebnisausgabe, das sich nach und nach mit Inhalt füllt. Je nach Leistungsfähigkeit des Computers kann die Analyse einige Zeit in Anspruch nehmen – der grüne Balken oben im Fenster zeigt den Fortschritt an:



Ist die Analyse beendet, werden oben die Untersuchungsdaten angezeigt (sofern vorhanden), darunter folgt die tabellarische Ausgabe der numerischen Ergebnisse:



Die erste Spalte listet die Parameter (14 akustische Stimmparameter (vgl. S. 6), 5 Pitchparameter und 2 Indizes). In der zweiten Spalte sind in grüner Farbe Normwerte aufgeführt (sofern vorhanden). Die Normwerte der Einzelparameter gelten für einen gehaltenen Vokal von 3 Sekunden Dauer, die AVQI- und ABI-Normen sind sprachspezifisch und gelten für die kombinierte Stimmprobe (MX). In der dritten und vierten Spalte sind die Ergebnisse

für die Vokal- (SV) und für die Text-Stimmprobe (CS) aufgeführt. Die fünfte Spalte zeigt die Ergebnisse für die kombinierte Stimmprobe (MX). Ergebnisse der Vokal-Stimmprobe, die außerhalb des Normbereichs liegen, werden in roter Farbe angezeigt (sofern Normwerte vorliegen), ebenso AVQI- und ABI-Werte oberhalb des Normbereichs.

Bei unvollständigen Stimmproben werden entsprechend nur Teilergebnisse ausgegeben. Im folgenden Beispiel lag nur eine Vokal-Stimmprobe vor mit weniger als 3 Sekunden Dauer:

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 diesem Fall werden alle numerischen Parameter der Vokal-Stimmprobe berechnet (3. Spalte), es werden jedoch keine Normwerte angegeben (2. Spalte), da diese nur für Vokal-Stimmproben mit 3 Sekunden Dauer gelten. Aufgrund der zu kurzen Vokal-Stimmprobe und der fehlenden Text-Stimmprobe werden AVQI und ABI nicht berechnet.

4.3.2 Stimmprofil

Entsprechen die Stimmproben der Spezifikation in Abschnitt 4.2 wird neben der textuellen Ergebnisausgabe im gleichen Fenster auch ein graphisches Stimmprofil erzeugt (s. Abb. 5). Das Stimmprofil zeigt kompakt auf einer Seite:

1. **Die Untersuchungsdaten:** Alle Daten, die zuvor in das Formular im linken Bereich des Hauptfensters eingetragen wurden, und das Aufnahmedatum der beiden Stimmproben.
2. **Ein Schmalband-Spektrogramm** des gehaltenen Vokals zur spektrografischen Stimmtypenklassifizierung ([6, 91]).
3. **Numerische Analyseergebnisse** des gehaltenen Vokals sowie AVQI- und ABI-Index; unauffällige Werte in schwarzer Farbe, abweichende Werte in roter Farbe; rechts daneben in grüner Farbe die Normwerte, sofern vorhanden.

4. Ein Kreisdiagramm, das die Analyseergebnisse der Vokal-Stimmprobe in 4 relevanten Dimensionen ([18], siehe auch Abschnitt 6.1) sowie die beiden Indizes AVQI und ABI, die auf der Basis der kombinierten Stimmprobe berechnet werden, mithilfe des intuitiven Ampelsystems illustriert. Der mittlere grüne Kreis repräsentiert den Normbereich. Abweichungen werden rot gekennzeichnet; je mehr ein Wert in einer Dimension von der Norm abweicht, desto weiter außen wird der Wert abgetragen. Liegt die Abweichung außerhalb des darstellbaren Bereichs, wird diese besonders starke Anomalie der Stimmqualität mit roten Pfeilen gekennzeichnet (s. Abb. 6).

Auch wenn nur eine Vokal-Stimmprobe mit 3 Sekunden Dauer vorliegt, wird ein graphisches Stimmprofil erzeugt, aufgrund der fehlenden Text-Stimmprobe können jedoch die Indizes AVQI und ABI nicht berechnet werden, so dass in diesem Fall nur die Einzelparameter berücksichtigt werden.

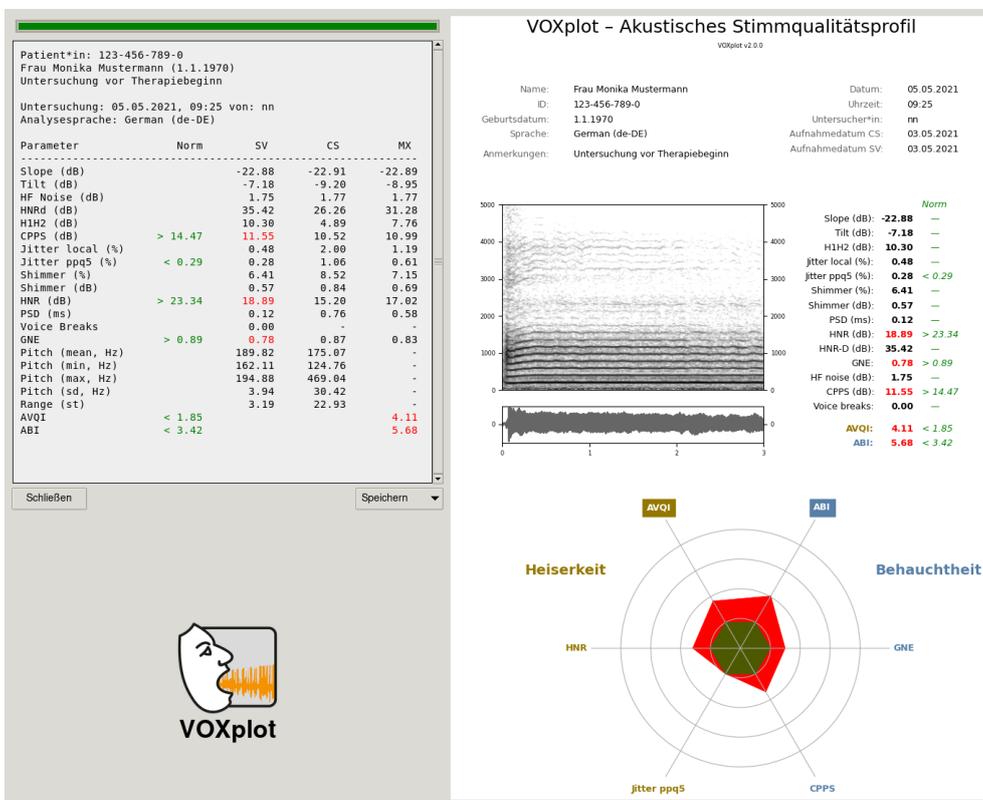


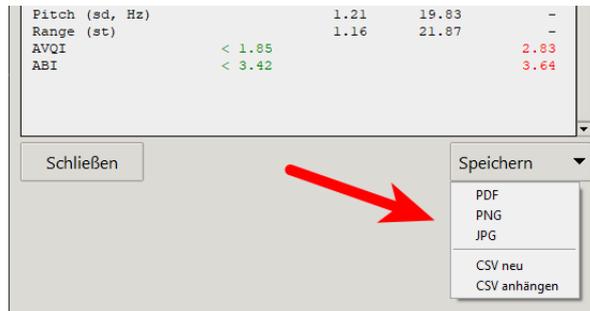
Abbildung 5: Die Ergebnisausgabe (links oben) und das graphische VOXplot Stimmprofil (rechts).



Abbildung 6: Norm-Abweichung außerhalb des darstellbaren Bereichs (hier der AVQI-Wert). Dies kennzeichnet eine besonders starke Anomalie der Stimmqualität hinsichtlich dieses Parameters.

4.3.3 Ergebnisse speichern

Zur Archivierung, zur Weitergabe oder zum Ausdrucken kann das graphische Stimmprofil sowohl als PDF-Datei als auch als Bild-Datei in den Formaten PNG oder JPG gespeichert werden. Das entsprechende **Speichern**-Menü finden Sie unterhalb der Ergebnisausgabe:



Sollen die Analyseergebnisse in anderer Form weiterverarbeitet werden (Excel, Calc, R, SPSS, GNU PSPP usw.), können die Untersuchungsdaten und die numerischen Ergebnisse auch im universellen CSV-Format gespeichert werden (in einer neu erzeugten CSV-Datei oder angehängt an eine bereits vorhandene CSV-Datei.). Das gewünschte CSV-Format (Feldtrennzeichen und Dezimalzeichen) sollte zuvor in den VOXplot-Einstellungen überprüft und ggf. spezifiziert werden (s. Abschnitt 5.4).

Mit der Option **CSV neu** wird eine neue CSV-Datei mit 2 Zeilen erzeugt: In der ersten Zeile steht der Header (Spaltentitel), die zweite Zeile beinhaltet die exportierten Daten. Mit der Option **CSV anhängen** kann eine bereits existierende CSV-Datei ausgewählt werden; die exportierten Daten werden dann dieser Datei in einer neuen Zeile hinzugefügt (ohne Header). Achtung: Die existierende Datei sollte in der ersten Zeile bereits einen Header besitzen und v.a. das gleiche CSV-Format aufweisen wie die exportierten Daten.

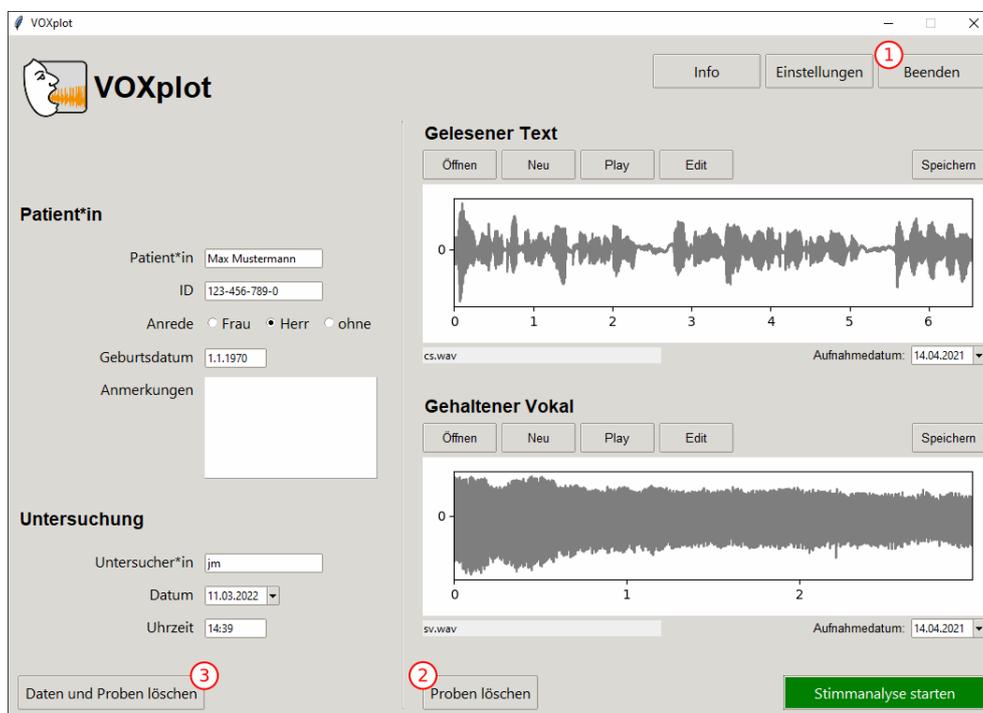
Insgesamt werden 70 Felder exportiert: Sämtliche Untersuchungsdaten, Aufnahmezeiten und Analysesprache sowie alle numerischen Analyseergebnisse. Die folgende Abbildung zeigt einen Ausschnitt (9 Felder bzw. Spalten) aus der resultierenden Tabelle nach dem Import der CSV-Datei in LibreOffice Calc:

The screenshot shows a LibreOffice Calc window with a CSV file imported. The table has 9 columns and 3 rows of data. The columns are labeled 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	1.508585678332246	1.5017547693333264	19.62434236506296	6.956582978592916	1.3304043887617845	18.703657707670008	5.206375606689481

Der Ausschnitt zeigt die 3 Parameter **hfno** (Hochfrequente Geräuschanteile), **hnrd** (HNR nach Dejonckere & Lebacq) und **h1h2** (Amplituden-Differenz zwischen 1. und 2. Harmonischen), jeweils für die Vokal-Stimmprobe (**sv_**), die Text-Stimmprobe (**cs_**) sowie die kombinierte Stimmprobe (**mx_**).

Nach dem Speichern der Ergebnisse kann die Ergebnisausgabe mit dem **Schließen**-Button geschlossen werden. Im Hauptfenster haben Sie nun die Wahl:



① VOXplot beenden

Achtung

Stimmproben, die innerhalb von VOXplot aufgenommen und noch nicht gespeichert wurden, gehen ohne Nachfrage verloren.

② Proben löschen

Möchten Sie von der untersuchten Person weitere Stimmproben öffnen oder aufnehmen, können mit **Proben löschen** die vorhandenen Stimmproben entfernt werden. Die Untersuchungsdaten im linken Bereich bleiben dabei erhalten (und können bei Bedarf editiert werden).

② Daten und Proben löschen

Soll eine andere Person untersucht werden, werden mit **Daten und Proben löschen** nicht nur vorhandene Stimmproben, sondern auch die Untersuchungsdaten gelöscht; Datum und Uhrzeit werden aktualisiert. Dies entspricht dem Zustand von VOXplot nach einem Neustart.

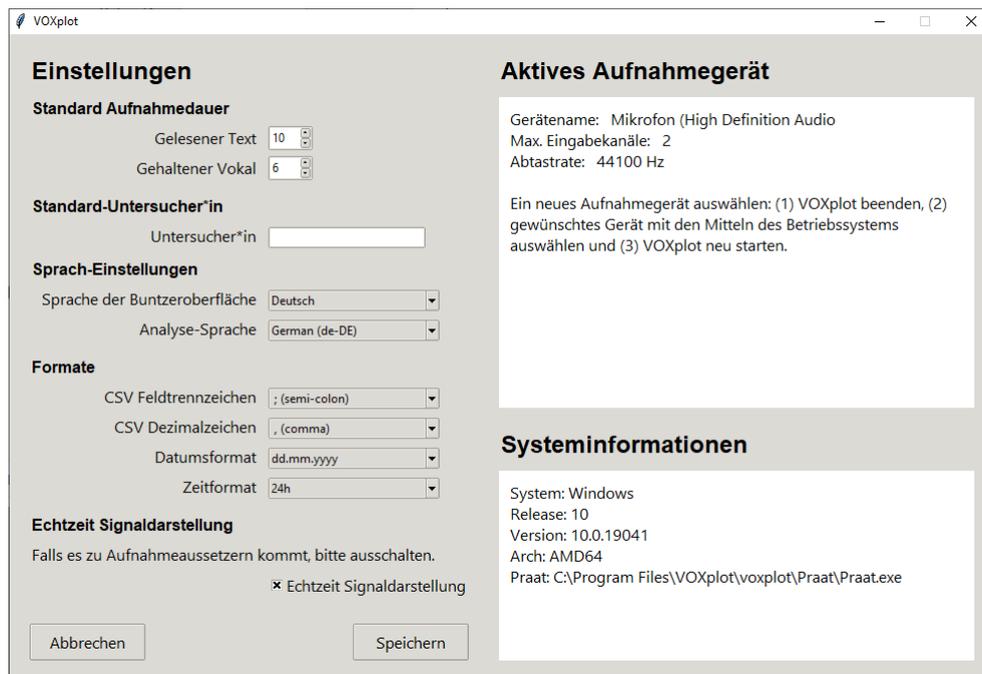


Abbildung 7: Die VOXplot-Einstellungen.

5 Einstellungen

Um die Einstellungen zu öffnen, klicken Sie im Hauptfenster auf **Einstellungen**. Das Einstellungsfenster zeigt rechts Informationen zum aktiven Aufnahmegerät und zum Computersystem an. Im linken Fensterbereich können Sie einige Einstellungen vornehmen, die das Verhalten von VOXplot steuern. Diese Einstellungen werden auf dem Computer in Ihrem Nutzerkonto gespeichert und bleiben erhalten, wenn VOXplot beendet wird.

5.1 Standard Aufnahmedauer

Die Aufnahmedauer (in Sekunden) kann getrennt nach Text und Vokal voreingestellt werden. Nach der hier eingestellten Aufnahmedauer bricht die jeweilige Aufnahme automatisch ab (wenn sie nicht zuvor manuell beendet wurde). Der höchste Wert für die Textaufnahme beträgt 30 Sekunden, für die Vokalaufnahme 15 Sekunden. Ab Werk sind 10 bzw. 6 Sekunden voreingestellt.

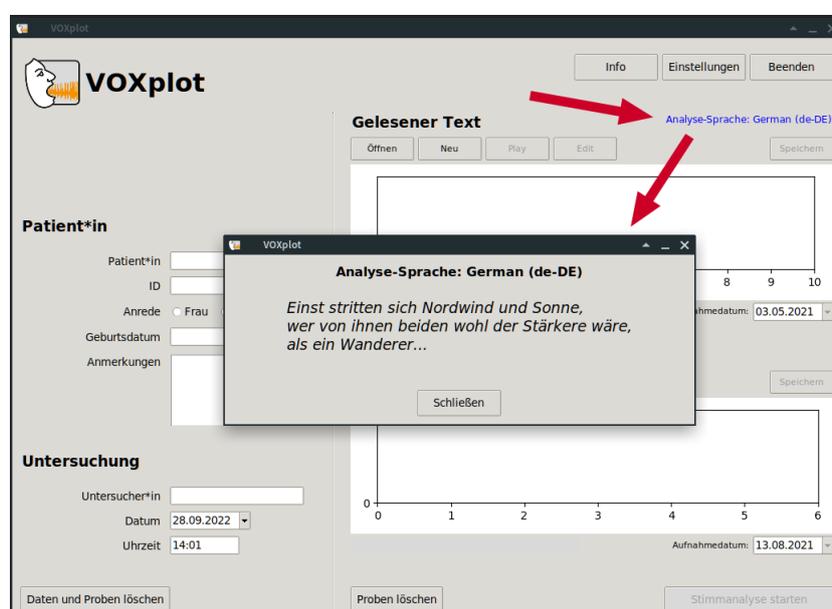
5.2 Standard-Untersucher*in

Werden Untersuchungen überwiegend von der selben Person durchgeführt, kann hier ein Name oder ein Kürzel eingetragen werden. Das Feld **Untersucher*in** im Hauptfenster wird dann automatisch mit diesem Namen/Kürzel befüllt.

5.3 Sprach-Einstellungen

Die Sprache der Benutzeroberfläche betrifft die Beschriftungen von Buttons und sonstigen Elementen der Programm-Oberfläche sowie einige Beschriftungen im Stimmprofil (s. Abb. 1). Diese Einstellung kann frei nach der Vorliebe der Untersucherin bzw. des Untersuchers gewählt werden.

Die Analyse-Sprache ist davon unabhängig und muss **unbedingt entsprechend der tatsächlich analysierten Sprache** gewählt werden. Die Stimme einer deutschsprachigen Person muss demnach mit der Einstellung *German* analysiert werden, die Stimme einer spanischsprachigen Person mit der Einstellung *Spanish* usw. Seit **VOXplot v1.3.0** wird die ausgewählte Analysesprache auch prominent im Hautfenster angezeigt; ein Klick auf die Sprache öffnet ein Fenster mit dem für diese Sprache vorgegebenen Lesetext:



5.4 Formate

VOXplot kann numerische Analyseergebnisse im universellen CSV-Format exportieren (s. Abschnitt 4.3.3). Mit den ersten beiden Einstellungen im Abschnitt *Formate* kann spezifiziert werden, mit welchem Zeichen die Felder in einem CSV-Datensatz getrennt werden sollen und welches Dezimalzeichen bei numerischen Daten verwendet werden soll. Standard ist das Komma als Trennzeichen (CSV = *comma separated values*) und der Dezimalpunkt. Je nachdem, wie die CSV-Daten weiterverarbeitet werden sollen, können hier andere Werte sinnvoll sein. Um die Daten z.B. in ein auf Deutsch voreingestelltes Excel zu importieren, ist das Semikolon als Trennzeichen und das Dezimalkomma besser geeignet.

Die Einstellung des Datums- und Zeitformats betrifft die Darstellung in den Feldern Datum bzw. Zeit im Hauptfenster sowie die Ausgabe im Stimmprofil.

5.5 Echtzeit Signaldarstellung

Ab Werk wird während einer laufenden Aufnahme der Signalverlauf (fast) in Echtzeit dargestellt. Auf weniger leistungsstarken Computern kann dies das System überlasten, die Folge sind Aussetzer in der Aufnahme. Wenn Sie solche Aussetzer beobachten, sollten Sie hier die Echtzeit-Darstellung deaktivieren. Zwar fehlt dann das direkte Feedback während der Aufnahme, aber die Signalaufzeichnung sollte unterbrechungsfrei durchlaufen.

5.6 Aktives Aufnahmegerät

An einen Computer können gleichzeitig mehrere Aufnahmegeräte angeschlossen sein, von denen aber normalerweise nur eines aktiv für eine Aufnahme genutzt wird. Beispielsweise kann man an ein Laptop zusätzlich zum eingebaute Mikrophon (das man im Rahmen der Stimmanalyse besser nicht nutzen sollte) per USB ein hochwertiges externes Mikrophon anschließen. Welches der angeschlossenen Geräte für Aufnahmen genutzt wird, kann im Betriebssystem (Windows, macOS, Linux) mit einschlägigen Audio-Mixer-Werkzeugen ausgewählt werden. **VOXplot** erkennt das aktive Aufnahmegerät und zeigt es im entsprechenden Abschnitt zusammen mit der Kanal-Anzahl und der eingestellten Abtastrate an. Sollte das angezeigte Gerät nicht das gewünschte Gerät sein, schließen Sie das Einstellungsfenster und beenden Sie **VOXplot**. Wählen Sie dann mit den Werkzeugen des Betriebssystems das gewünschte Gerät aus und starten Sie **VOXplot** neu.

6 Akustische Stimmparameter und ihre Interpretation

6.1 Vorbemerkungen

Die Berechnung akustischer Stimmparameter hängt von zahlreichen Faktoren ab, was ein Problem für die Interpretation und Vergleichbarkeit der Werte darstellt. Die Einflussfaktoren reichen vom verwendeten Mikrophon, über die restliche Aufnahme-Hardware, die Raumakustik, die gewählte Abtastrate bis hin – und dies ist der Faktor mit dem größten Einfluss – zu den Algorithmen, mit denen z.B. Stimmhaftigkeit im akustischen Signal detektiert oder Jitter bestimmt wird [5, 26, 33–36, 82, 93, 96, 97].

Daher sind z.B. Norm- bzw. Grenzwerte für bestimmte Parameter mit einer gewissen Vorsicht zu genießen. Schon die Verwendung eines anderen Mikrophons oder einer anderen Abtastrate kann die Aussagekraft solcher Normwerte reduzieren. Praktisch unbrauchbar werden Normwerte dann, wenn die entsprechenden Parameter mit unterschiedlichen Algorithmen berechnet werden. Dies trifft z.B. auf **VOXplot/Praat** im Vergleich mit dem kommerziellen MDVP zu [2, 24, 70, 80]. Die zugrundeliegenden Algorithmen unterscheiden sich erheblich zwischen den beiden Programmen, so dass die MDVP-Normwerte, wenn überhaupt, nur als ganz grober Anhaltspunkt bei der Interpretation einer **VOXplot**-Stimmanalyse dienen können und umgekehrt. Eine Erklärung dafür, wie und weshalb sich die Algorithmen zwischen **VOXplot/Praat** und MDVP unterscheiden und welche Konsequenzen dies für mit Praat berechnete Werte im Vergleich zu MDVP-Werten hat, ist u.a. in der Praat-Hilfe nachzulesen.

VOXplot stellt einige eigene Normwerte zur Verfügung. Die Grenzwerte für die multiparametrischen Indizes AVQI und ABI sind sprachspezifisch und werden nur in **VOXplot** integriert, wenn entsprechende wissenschaftlich fundierte, sprachspezifische Studien vorliegen [1, 9, 15, 17, 32, 38–40, 50, 51, 53, 55, 57, 58]. Darüberhinaus werden seit **VOXplot** v2.0 nur noch Normwerte für vier akustische Einzelparameter der (nicht sprachspezifischen) Vokalstimmprobe angegeben: HNR, Jitter ppq5, CPPS und GNE. Wie in einer aktuellen Studie auf der Basis von 218 Stimmproben gezeigt werden konnte, können aus den untersuchten 13 Parametern⁵ nur diese vier Parameter valide mit der perzeptiven Beurteilung von Heiserkeit bzw. Behauchtheit assoziiert werden [18]. Kriterien für die Zuordnung eines Parameters zu einer der beiden Stimmanomalien waren: (a) eine hohe Korrelation ($Correlation > 0,70$ und $A_{ROC} > 0,70$) mit Heiserkeit oder Behauchtheit und (b) eine signifikante Differenz bei Korrelation oder ROC -Statistik bzgl. Heiserkeit und Behauchtheit. Anders ausgedrückt haben nur diejenigen Parameter die Einschlusskriterien erfüllt, die stark mit der einen Stimmanomalie und zugleich signifikant weniger mit der anderen Stimmanomalie assoziiert waren. Für die Heiserkeit trifft dies auf die beiden Parameter *Harmonics-to-Noise Ratio* (**HNR**) und den Pitch-Perturbationsquotienten mit einem Glättungsfaktor von 5 Perioden (**Jitter ppq5**) zu. Mit Behauchtheit assoziiert sind die *Smoothed Cepstral Peak Prominence* (**CPPS**) und die *Glottal-to-Noise Excitation Ratio* (**GNE**).

⁵Bei den 13 Parametern handelt es sich um die **VOXplot**-Parameter, die in den folgenden Abschnitten erläutert werden, mit Ausnahme des unspezifischen Parameters *Voice breaks*.

Achtung

Norm- bzw. Grenzwerte, die in diesem Handbuch oder in der VOXplot-Ergebnisausgabe angegeben werden, sind nicht übertragbar. Sie gelten nur im Zusammenhang mit Untersuchungsergebnissen, die mit VOXplot ermittelt wurden.

Das Normwerteproblem wird entschärft, wenn VOXplot für Verlaufsuntersuchungen z.B. im Rahmen der Dokumentation bzw. Evaluation einer Stimmtherapie eingesetzt wird und die in Kapitel 8 formulierten Empfehlungen zum Aufnahme-Setting beachtet werden. In diesem Fall werden die akustischen Stimmparameter von einer Person zu verschiedenen Zeitpunkten erhoben und miteinander anstatt mit Normwerten verglichen. Auf diese Weise können positive oder negative Veränderungen der Stimme objektiv dokumentiert werden. Die Entscheidung, ob sich ein Parameter zum Positiven oder zum Negativen verändert hat, ist i.d.R. einfach zu treffen. Verlässlich sind aber auch solche Vergleiche nur, wenn das Aufnahme-Setting (Mikrophon, Mikrophonabstand, sonstige Aufnahme-Hardware, Umgebung etc.) sowie das Erhebungsverfahren über die verschiedenen Messungen hinweg möglichst identisch gehalten werden.

Generell gilt, dass Parameter, vor deren Berechnung eine Analyse der Periodizität des Zeitsignals durchgeführt werden muss (z.B. Jitter, Shimmer oder PSD), problematischer sind als spektrale Parameter (z.B. Slope, Tilt oder CPPS). Für die Periodizitätsanalyse kann auf zahlreiche Algorithmen zurückgegriffen werden, wobei gilt: Alle Algorithmen sind fehlbar und die Fehleranzahl steigt – bis hin zum vollständigen Versagen – je problematischer das Signal. Als problematisch in dieser Hinsicht müssen alle Signale gelten, die nicht mit modaler Stimme erzeugt wurden, also gerade diejenigen Signale, die in der Stimmanalyse häufig von besonderem Interesse sind.

Für spektrale Parameter gilt diese Voraussetzung nicht, da deren Berechnung erfolgt, nachdem das Zeitsignal mittels spektraler Analyse in die Frequenzdomäne transferiert wurde (Fourier-Transformation). Insofern sind spektrale Parameter robuster als beispielsweise Perturbationsparameter. Ein weiterer Vorteil besteht darin, dass neben gehaltenen Vokalen auch natürliche, fortlaufende Äußerungen sinnvoll analysiert werden können. Dies wird möglich durch die Verwendung eines *Long-Term Average Spectrum* (LTAS). Bei dieser Analyse-Methode werden die spektralen Eigenschaften eines Signals über einen längeren Abschnitt integriert und als Durchschnittsspektrum berechnet.

6.2 Spektrale Neigung

[Slope, Tilt]

Unter spektraler Neigung versteht man die globale Energieverteilung im Leistungsspektrum. Spektren stimmhafter Segmente weisen in den tieferen Frequenzbändern mehr Energie auf als in den höheren Frequenzbändern; bei modaler Stimme beträgt der Energieabfall im Quellspektrum typischerweise rund 12 dB/Oktave.

Insbesondere bei behauchten Stimmen, also verlängerter Offen-Phase, ist dieser Effekt verstärkt, d.h. der Energieabfall in den höheren Frequenzbändern ist ausgeprägter als bei modaler Stimme, das Spektrum ist folglich stärker geneigt. Bei verlängerter Geschlossen-Phase kann es dagegen zu einer Anhebung der Energie in höheren Frequenzbereichen kommen, das Spektrum ist weniger stark geneigt.

In der Literatur werden zahlreiche Methoden beschrieben, die spektrale Neigung zu messen.⁶ **VOXplot** nutzt zwei davon: (1) Die globale Energieverteilung im Long-Term Average Spectrum (*Slope*) und (2) das Gefälle der Regressionslinie durch ein LTAS (*Tilt*). Beide Parameter gehen auch in den *Acoustic Voice Quality Index* (AVQI) ein, der im Kapitel 7.1 vorgestellt wird.

Der *Slope* wird berechnet als Energie-Differenz zwischen zwei frei wählbaren Frequenzbändern. Um mit dem AVQI-Algorithmus kompatibel zu sein, nutzt **VOXplot** als tiefes Frequenzband 0 Hz bis 1000 Hz, das hohe Frequenzband reicht von 1 kHz bis 10 kHz. Nimmt die Energie zu den höheren Frequenzen hin ab, erhält man einen negativen *Slope*-Wert, d.h. bei menschlichen Stimmen erhält man immer einen *Slope*-Wert < 0 . Stärker geneigte Spektren ergeben kleinere Werte (z.B. -30 dB) als weniger geneigte Spektren (z.B. -10 dB). Als sehr groben (!) Anhaltspunkt kann man bei modalen Stimmen einen *Slope*-Wert zwischen -10 und -20 dB erwarten; bisweilen findet man aber auch bei gesunden Stimmen Werte *außerhalb* dieses Bereichs; desgleichen findet man bei pathologischen Stimmen Werte *innerhalb* dieses Bereichs. Für sich genommen besitzt folglich ein einzelner *Slope*-Wert kaum Aussagekraft. Vergleicht man jedoch zwei Messungen, kann man z.B. bei einer Steigerung von -35 dB auf -20 dB objektiv von einer Verbesserung sprechen.

Der *Tilt*-Parameter beschreibt das Gefälle der Regressionslinie durch das LTAS. Wie beim *Slope*-Parameter gilt: Je stärker das Gefälle der Linie desto kleiner der *Tilt*-Wert (starkes Gefälle z.B. -14 , schwaches Gefälle z.B. -6).

6.3 Amplitudendifferenz H1-H2

[H1H2]

Der Stimmtton entsteht durch mehr oder weniger regelmäßiges Öffnen und Schließen der Glottis. Jeder Phonationszyklus besteht aus einer Offen- und einer Geschlossen-Phase der Glottis. Der zeitliche Anteil dieser Phasen am Gesamtzyklus korreliert mit der Stimmqualität: Bei modalen Stimmen ist die Offen-Phase meist etwas länger als die Geschlossen-Phase. Behauchte Stimmen weisen sehr lange Offen-Phasen und kurze Geschlossen-Phasen auf, bei Laryngalisierung (*creaky voice*) bzw. gepressten Stimmen ist dagegen häufig die Geschlossenphase verlängert, was meist zum Eindruck einer rauhen Stimme führt. Liegt eine EGG-Messung⁷ vor, können die Phasendauern direkt und sehr genau im laryngalen Erregungssignal (Lx-Signal) gemessen werden; dies ist in Abb. 8 dargestellt.

⁶Auch die im nächsten Abschnitt diskutierte Amplitudendifferenz zwischen H1 und H2 kann als Maß für die spektrale Neigung gelten.

⁷Die Elektroglottographie (EGG) ist eine nicht-invasive Methode, um die glottale Aktivität zu messen. Zwischen zwei Oberflächenelektroden, die links und rechts des Schildknorpels fixiert werden, wird die elektrische Leitfähigkeit

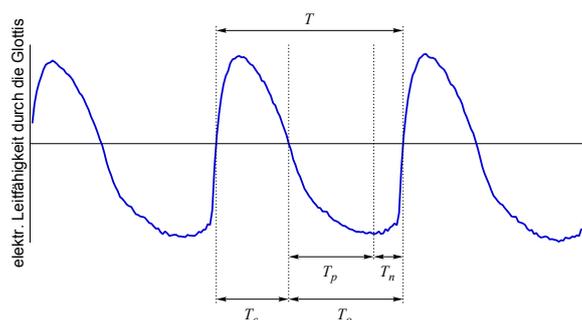


Abbildung 8: Das EGG-generierte laryngale Erregungssignal (Lx-Signal). T = Periodendauer einer Stimmlippenschwingung (Phonationszyklus); T_o = Offen-Phase; T_c = Geschlossen-Phase; T_p = zunehmender Luftstrom; T_n = abnehmender Luftstrom.

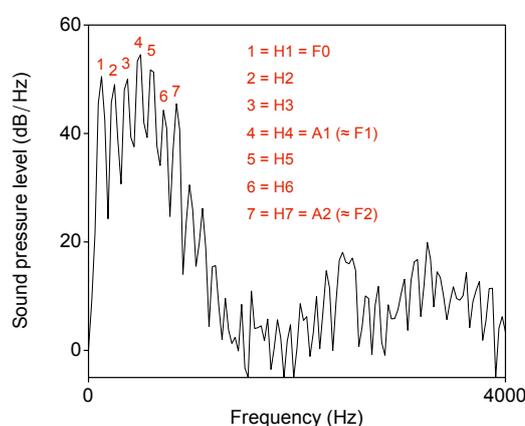


Abbildung 9: Schmalbandspektrum eines Vokals /a:/; die ersten sieben Harmonischen sind durch die Ziffern 1-7 gekennzeichnet.

Ohne EGG sind wir auf die (weniger zuverlässige) Messung akustischer Korrelate angewiesen. Das stabilste akustische Korrelat für das Verhältnis zwischen Offen- und Geschlossen-Phase ist die Amplitudendifferenz zwischen der ersten und zweiten Harmonischen im Spektrum. Abbildung 9 zeigt das Schmalbandspektrum eines Vokals /a:/; die ersten 7 Harmonischen sind gekennzeichnet und durchnummeriert. H1 entspricht der Grundfrequenz (F_0). Bei diesem Vokal und dieser Sprecherin wird H4 und H7 durch den ersten bzw. zweiten Formanten verstärkt, daher die Bezeichnung A1 bzw. A2.⁸ Selbstverständlich betrifft dieser Verstärkungseffekt bei anderen Vokalen und/oder anderen Sprechern auch andere Harmonische als H4/H7.

keit gemessen. Bei geschlossener Glottis ist die Leitfähigkeit groß, bei offener Glottis ist sie wegen des Luftspaltes zwischen den Stimmlippen gering.

⁸A1 bezeichnet die Amplitude der Harmonischen in der Nähe des ersten Formanten; A2 bezeichnet entsprechend die Amplitude der Harmonischen in der Nähe des zweiten Formanten.

H2 hat in Abb. 9 eine etwas geringere Amplitude (einen 'niedrigeren Gipfel') als H1, d.h. wenn man die Amplitudendifferenz $H1-H2$ berechnet, erhält man einen kleinen positiven Wert. Ein größerer Offen-Quotient ($\frac{T_o}{T_c}$), d.h. eine gelängte Offen-Phase, hat eine geringere H2-Amplitude zur Folge, entsprechend ergibt $H1-H2$ einen höheren positiven Wert (> 4). Bei kleinerem Offen-Quotient, d.h. bei einer gelängten Geschlossen-Phase, steigt die H2-Amplitude an. Steigt sie stark an und übertrifft die H1-Amplitude, ergibt $H1-H2$ einen negativen Wert (< -2).

Die folgende Darstellung des $H1-H2$ -Wertebereichs dient nur der groben Orientierung, bei den 'Grenzwerten' handelt es sich nicht um ordentlich erhobene Normdaten!

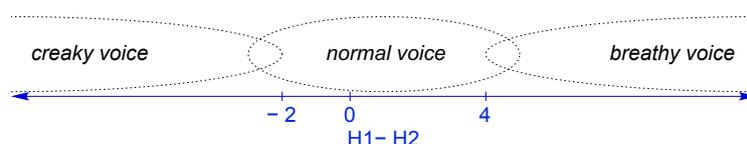


Abbildung 10 zeigt ein Beispiel mit normaler Stimme ($H1-H2 = 1.3$) und zwei Beispiele mit pathologisch veränderten Stimmen: Hochgradig behaucht ($H1-H2 = 12.2$) und mittelgradig rau ($H1-H2 = -2.2$).

Achtung: Der Parameter $H1-H2$ ist sehr anfällig für Interferenzen, die nichts mit Stimmqualität zu tun haben. Der erste Einflussfaktor, dem man relativ einfach aus dem Weg gehen kann, ist der erste Vokalformant. Bei Vokalen mit tiefem erstem Formanten (das sind die geschlossenen Vokale) besteht die Gefahr, dass der erste Formant die Amplitude der tiefen Harmonischen anhebt, wodurch die Berechnung der Amplitudendifferenz 'verfälscht' wird, d.h. das Ergebnis kann nicht mehr hinsichtlich Stimmqualität interpretiert werden. Als Grundlage für die Messung eignen sich daher am besten offene Vokale mit hohem erstem Formanten; am besten geeignet ist der Vokal /a/.

Der zweite Faktor, der Einfluss auf die Amplitude der tiefen Harmonischen ausübt, ist Nasalität. Ist ein Vokal ganz oder teilweise nasalisiert, wird die erste Resonanzfrequenz des Nasenraumes (zwischen 200 und 300 Hz) meist die Amplitude der ersten (hohe Stimmen) oder zweiten Harmonischen (tiefe Stimmen) anheben. Die Konsequenz ist die selbe wie oben: Das Ergebnis der Amplitudendifferenz zwischen H1 und H2 kann nicht mehr hinsichtlich Stimmqualität interpretiert werden [90]. Zwar sieht die deutsche Phonologie keine Vokalnasalierung vor, dennoch findet man auch im Deutschen häufig zumindest teilweise nasalisierte Vokale als Folge koartikulatorischer Prozesse im Kontext von Nasalkonsonanten. Desweiteren kann diese Messung offensichtlich bei dysarthrischen Sprechern mit Hypernasalität nicht sinnvoll angewendet werden. Zuletzt sei noch auf das Problem der intrinsischen Nasalität hingewiesen: Je offener der Vokal desto tiefer das Velum [22]. D.h. offene Vokale, die aufgrund ihres hohen ersten Formanten eben noch als ideale Messgrundlage erschienen, haben von allen Vokalen die höchste intrinsische Nasalität und müssen daher eher als problematisch gelten.

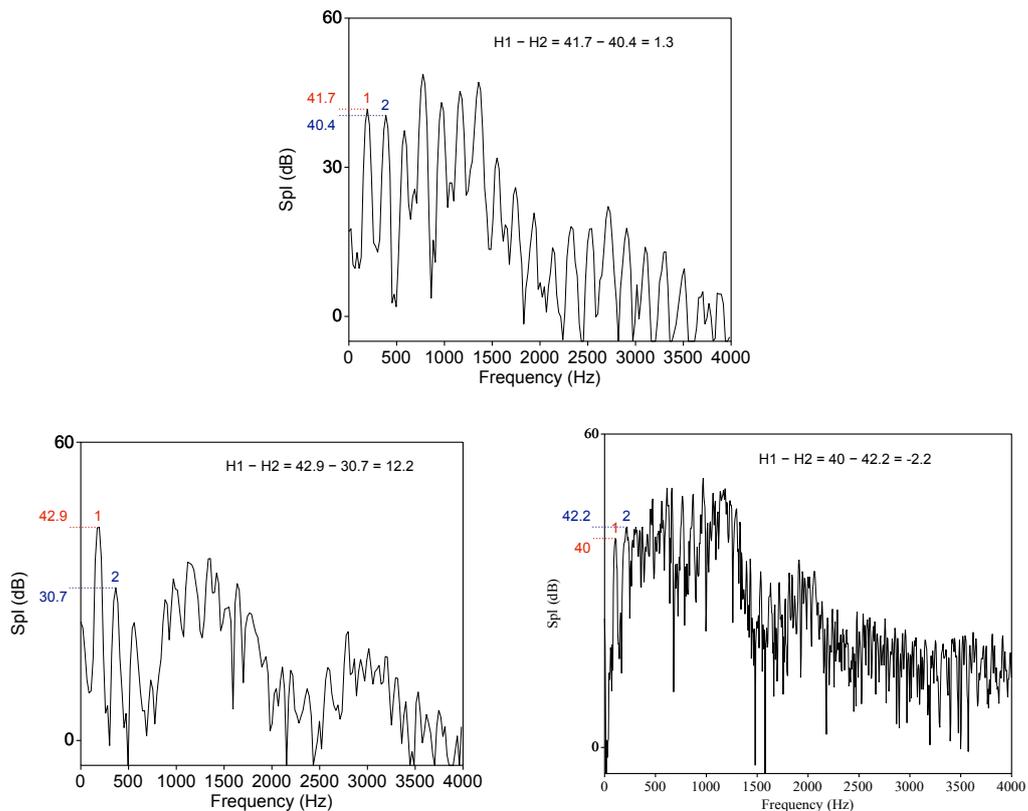


Abbildung 10: Spektren von 3 verschiedenen Sprecherinnen bzw. Sprechern: Normale Stimme (oben), hochgradig behauchte Stimme (unten links) und mittelgradig raue Stimme (unten rechts).

Trotzdem wird meist der Vokal /a/ verwendet. Man geht davon aus, dass der handfeste Vorteil des hohen ersten Vokalformanten gegenüber dem potentiellen Nachteil der intrinsichen Nasalität überwiegt. Selbstverständlich muss die Messung verworfen werden, wenn ein Vokal der Nasalierung verdächtig ist.

Zum Abschluss dieser Diskussion zeigt Abb. 11 am Beispiel eines Vokals /a:/ vor einem Nasalkonsonanten (aus dem Wort *Fahne*) den Einfluss der antizipatorischen Nasalierung auf das Verhältnis zwischen H1 und H2. Die Amplitudendifferenz am Anfang, im nicht-nasalierten Teil des Vokals, beträgt für eine modale Stimme typische 2.1 dB. Am Ende des Vokals, wenn sich das Velum bereits senkt, wird die Amplitude von H1 vom ersten Nasalformanten angehoben, während die Amplitude von H2 gedämpft wird (vermutlich unter dem Einfluss eines nasalen 'Antiformanten'); dies führt zu einer Amplitudendifferenz von 13.9 dB – an und für sich ein typischer Wert für eine stark behauchte Stimme, der hier jedoch ausschließlich auf die Nasalierung zurückzuführen ist.

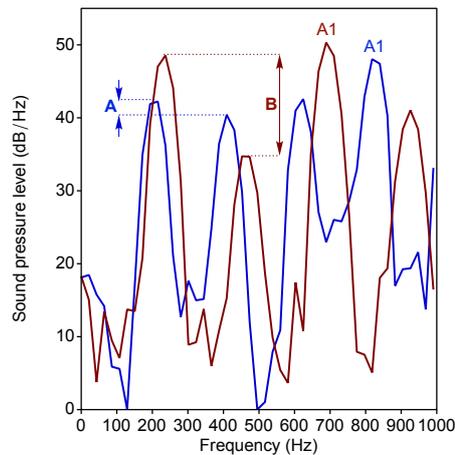


Abbildung 11: Spektrum am Anfang (blau) und am Ende (rot) des Vokals /a:/ im Wort Fahne (weibliche Sprecherin); am Anfang, im oralen Teil des Vokals beträgt $H1-H2 = 2.1$ (A); im nasalisierten Teil am Ende des Vokals ist die $H1$ -Amplitude angehoben und die $H2$ -Amplitude gedämpft, $H1-H2 = 13.9$ (B); weiter ist zu beobachten, dass die Grundfrequenz zum Vokalende ansteigt, während die Frequenz des ersten Vokalformanten in der VC-Transition fällt.

6.4 Perturbation

[Jitter, Shimmer, PSD]

Phoniertes Sprechen führt zu einem quasiperiodischen akustischen Signal. Quasiperiodisch bedeutet, das Signal ist periodisch genug, so dass ein menschlicher Hörer einen Klang wahrnehmen und eine Grundfrequenz (mithin eine Stimmtonhöhe) identifizieren kann, das Signal ist jedoch nicht periodisch in einem streng mathematischen Sinne, d.h. die aufeinanderfolgenden Perioden sind nicht exakt identisch. Abweichungen zwischen benachbarten Perioden (Perturbation) betreffen sowohl die Periodendauer und damit die Frequenz als auch die Amplitude. Die Grundfrequenz-Perturbation wird *Jitter* genannt, die Amplituden-Perturbation *Shimmer* (vgl. Abb. 12). In geringem Ausmaß sind Jitter und Shimmer also normale Erscheinungen der menschlichen Stimme. Auch leicht erhöhte Jitter- und Shimmer-Werte z.B. bei starken Rauchern sind auditiv oft noch unauffällig. Erst stark erhöhte Werte, wie sie bei verschiedenen Stimmstörungen zu beobachten sind, führen zum Eindruck einer heiseren Stimme.

Generell gilt für alle Jitter- und Shimmer-Parameter: Je kleiner der Wert desto besser. Wenn sich also die Stimme eines Patienten bei einer Verlaufsuntersuchung von *Jitter local* = 3,27% zum Zeitpunkt 1 hin zu *Jitter local* = 0,73% zum Zeitpunkt 2 verändert, kann objektiv von einer Verbesserung ausgegangen werden. Gesunde Stimmen weisen i.d.R. Jitter-local-Werte < 0,5% und Shimmer-Werte < 3% auf.⁹ Zu beachten ist, dass bei natürlichem, fortlaufendem

⁹Aufgrund unterschiedlicher Algorithmen zur Periodendetektion sind die Jitter-Messungen und infolgedessen die Jitter-Grenzwerte von VOXplot/Praat (0,5%) und MDVP (1,04%) nicht vergleichbar [2, 24, 70, 80].

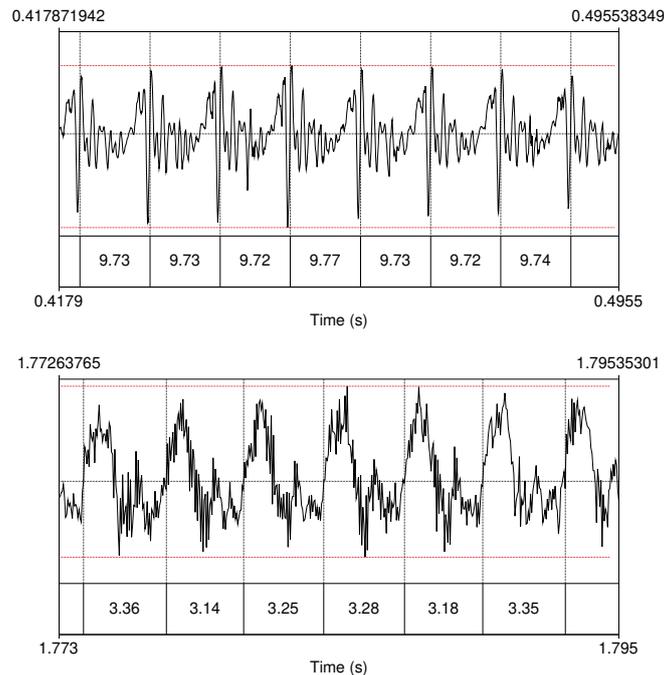


Abbildung 12: Ausschnitt aus einem gehaltenen Vokal: Normale Stimme (oben) und heißere Stimme (unten). Unterhalb des Signals ist jeweils die Periodendauer in Millisekunden abgetragen. Rot markiert ist die maximale Amplitudenvariation innerhalb des jeweiligen Ausschnitts. Die maximale Abweichung in der Periodendauer (Jitter) beträgt im oberen Ausschnitt 0,5% (9,72 ms vs. 9,77 ms) gegenüber 6,6% (3,14 ms vs. 3,36 ms) im unteren Ausschnitt. Abweichungen im Amplitudenbereich von Periode zu Periode (Shimmer) sind mit bloßem Auge erkennbar. (Sehr ausgeprägt und daher ebenfalls gut erkennbar sind die hochfrequenten, geräuschhaften Signalanteile im unteren Ausschnitt; vgl. Abschnitt 6.5.)

Sprechen meist höhere Jitter- und Shimmer-Werte resultieren als bei gehaltener Phonation. Die Orientierungswerte von 0,5% bzw. 3% gelten daher nur für gehaltene Phonation.

Jitter kann mit unterschiedlichen Algorithmen ermittelt werden. Sehr häufig wird *Jitter local* verwendet, die durchschnittliche Differenz zwischen zwei benachbarten Perioden, geteilt durch die durchschnittliche Periode. **VOXplot** berechnet zusätzlich den 5-point Period Perturbation Quotient (*Jitter ppq5*), der über 5 benachbarte Perioden glättet, also etwas weniger lokal arbeitet und somit als robustere Messung angesehen werden kann. In [18] wurde *Jitter ppq5* als valider Marker für Heiserkeit mit einem Grenzwert von 0,29% identifiziert. Interessierte finden mathematische Details in der Praat-Hilfe.

Period Standard Deviation (PSD) PSD beschreibt die Standardabweichung der Periodendauer in einem Signalabschnitt und gibt somit – ähnlich wie Jitter – Auskunft über die Frequenz-Perturbation eines quasiperiodischen Signals. Eine regelmäßige, periodische Phonation hat eine geringe Standardabweichung der Periodendauer zur Folge, deshalb gilt auch hier: Je kleiner der Wert desto besser.

6.5 Harmonizität

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

Unter dem Stichwort Harmonizität wird das Verhältnis zwischen harmonischen und nicht-harmonischen (geräuschhaften) Signalanteilen im quasiperiodischen Anregungssignal bewertet. Ein geringes Maß an Geräuschanteilen ist auch bei der gesunden Stimme zu beobachten. Ausgeprägte Geräuschanteile sind dagegen ein typischer Indikator für Behauchtheit bzw. generell Heiserkeit. **VOXplot** liefert drei Maße zur Beurteilung der Harmonizität: *HNR*, *HNR-D* und *GNE*.

HNR

Die *Harmonics-to-Noise Ratio* (HNR) gibt das Verhältnis von harmonischen bzw. periodischen Anteilen zu Geräuschanteilen im Anregungssignal an. Eine HNR von 20 dB bedeutet, dass 99% der Signalenergie auf periodische Signalanteile zurückzuführen sind und 1% auf Geräuschanteile. Eine HNR von 0% bedeutet, dass die Signalenergie zu gleichen Teilen auf periodische und geräuschhafte Anteile zurückzuführen ist. Interpretation: Je größer der Wert, desto besser. Bei gesunden Stimmen werden i.d.R. Werte über 20 dB erreicht, der **VOXplot**-Grenzwert liegt bei 23,34 dB (gehaltene Phonation) [18].

Bei irregulären Stimmlippenschwingungen, wie sie bei zahlreichen Stimmstörungen zu beobachten sind, ist der klassische HNR-Algorithmus schnell überfordert, da er auf Korrelationsanalysen in der Zeitdomäne beruht. In solchen Fällen bieten die folgenden beiden Parameter Vorteile, insbesondere die *Glottal-to-Noise-Excitation Ratio*.

HNR-D

Dejonckere und Lebacqz haben 1987 eine Methode publiziert, um die *Harmonics-to-Noise Ratio* im Spektrum zu ermitteln [31]. Hierfür wird die harmonische Struktur im Frequenzband zwischen 500 Hz und 1500 Hz basierend auf der cepstral ermittelten Grundfrequenz analysiert. Das analysierte Frequenzband entspricht dabei in etwa dem Bereich zwischen dem ersten und zweiten Formanten eines gehaltenen /a/. Auch hier gilt: Je größer der Wert, desto besser; gesunde Stimmen erreichen i.d.R. Werte deutlich über 20 dB.

Glottal-to-Noise-Excitation Ratio (GNE)

Geräuschanteile im Anregungssignal können auf unterschiedliche Ursachen zurückgeführt werden: (1) Perturbationsrauschen (s. Abschnitt 6.4) und (2) additives Rauschen. Perturbationsrauschen wird von irregulären Stimmlippenschwingungen verursacht und kann mit Hilfe von Jitter- und Shimmer-Messungen quantifiziert werden. Harmonizitätsmaße wollen dagegen den Anteil additiven Rauschens im Signal quantifizieren. Additives Rauschen wird durch Turbulenzen verursacht, wie sie z.B. bei unvollständig schließenden Stimmlippen entstehen, und korreliert daher mit Behauchtheit (im Gegensatz zu Jitter/Shimmer). Wie bereits erwähnt, ist die klassische HNR-Analyse anfällig bei irregulären Stimmlippenschwingungen (d.h. bei erhöhten Jitter- bzw. Shimmer-Werten) und daher nicht gut geeignet,

die beiden Rauschursachen analytisch zu trennen. Die *Glottal-to-Noise-Excitation Ratio* erweist sich dagegen als weitgehend unabhängig von Perturbationsrauschen und somit als robustes Maß für additives Rauschen auch bei ausgeprägter Irregularität einer Stimme. Das Berechnungsverfahren wird in [76] beschrieben. Je größer der Wert, desto besser. Der VOXplot-Grenzwert liegt bei 0,89 (gehaltene Phonation) [18].

Hochfrequenzrauschen (HF Noise)

HF Noise ist ein Quotient, der das Verhältnis zwischen der relativen akustischen Energie unterhalb einer Referenzfrequenz mit der relativen akustischen Energie oberhalb der gleichen Referenzfrequenz beschreibt.[30] VOXplot nutzt Long-Term-Average-Spektren (LTAS) mit einer Referenzfrequenz von 6 kHz zur Berechnung des Quotienten. Heißere Stimmen mit hohem additiven Rauschanteil weisen mehr Energie im Frequenzband oberhalb 6 kHz auf (→ kleinerer Quotient) als modale Stimmen (→ größerer Quotient). Je größer der Wert, desto besser. Der VOXplot-Grenzwert liegt bei 2,28 dB (gehaltene Phonation).

6.6 Stimmabbrüche

[Voice Breaks]

Degree of voice breaks ist ein Maß für den zeitlichen Anteil nicht-phonierte Abschnitte an der Gesamtdauer des analysierten Abschnitts. Werte größer als 0 können als Abweichung von der Norm gelten. Dieser Wert wird nur berechnet, wenn es sich bei der Stimmprobe um gehaltene Phonation handelt. Bei fortlaufendem Sprechen kommt es naturgemäß ständig zu 'Stimmabbrüchen' (stimmlose Segmente, Sprechpausen etc.), insofern ergibt der Parameter in diesem Fall keinen Sinn.

6.7 Cepstraler Gipfel

[CPPS]

Seit die Bedeutung der *Smoothed Cepstral Peak Prominence* (CPPS) für die akustische Stimm-analyse von James Hillenbrand und Kollegen Mitte der 1990er Jahre postuliert wurde [47, 48], konnten zahlreiche Studien zeigen, dass CPPS einerseits sehr robust und andererseits tatsächlich von allen bisher besprochenen Parametern derjenige mit der größten Relevanz ist, d.h. derjenige, der z.B. am besten mit auditiven Stimmqualitätsurteilen – insbesondere in der Behauchtheitsdimension – korreliert. Auch zur Bestimmung des Schweregrades einer Dysphonie eignet sich CPPS als objektives Maß [4, 16, 27, 42, 45, 63, 64, 71, 73, 78, 83, 88, 89].

CPPS kann als Maß für 'spektrales Rauschen' verstanden werden und basiert auf der Berechnung eines sog. Cepstrums. Es handelt sich dabei um eine inverse Fourier-Transformation eines logarithmierten Spektrums, d.h. – vereinfacht gesagt – ein Zeitsignal wird zunächst einer Spektralanalyse unterzogen und das Ergebnis wird dann einer weiteren Spektralanalyse unterzogen. Die Erfinder der Methode haben die Bezeichnung *Cepstrum* aus dem Wort *Spectrum* abgeleitet – die ersten vier Buchstaben sind vertauscht [25]. Entsprechend finden wir auf der x-Achse nicht die *Frequency* abgetragen sondern die *Quefrequency*, die Entsprechung

der Harmonischen (*Harmonics*) heißt *Rahmonics* und das Ganze ist keine *analysis* sondern eine *alanysis*...

Bei der CPPS-Messung macht man sich die folgende Gesetzmäßigkeit zu Nutze: Je ausgeprägter die Periodizität eines Zeitsignals desto ausgeprägter die harmonische Struktur des resultierenden Spektrums und desto prominenter der erste Gipfel (Rahmonische) des resultierenden Cepstrums. Die Amplitude dieses cepstralen Gipfels reflektiert somit die Intensität der harmonischen Struktur des Amplitudenspektrums: Eine ausgeprägte harmonische Struktur hat einen prominenten cepstralen Gipfel zur Folge; eine schwach ausgeprägte harmonische Struktur mit starkem 'spektralen Rauschen' hat ein flaches Cepstrum bzw. einen cepstralen Gipfel mit geringer Amplitude zur Folge [41].

Gemessen wird die Prominenz des cepstralen Gipfels üblicherweise in einem geglätteten Cepstrum, daher die Bezeichnung: *Smoothed Cepstral Peak Prominence*. Abbildung 13 zeigt zum Vergleich jeweils das Spektrum, das daraus abgeleitete Cepstrum sowie das geglättete Cepstrum eines gehaltenen Vokals mit modaler Stimme (links) bzw. mit hochgradig heiserer Stimme (rechts; Zustand nach Chordektomie mit Rezidiv-Granulationspolyp in vorderer Kommissur). Anhand der an der Grundfrequenz ausgerichteten harmonischen Gitterlinien im Spektrum wird der Unterschied im Grad der Periodizität zwischen den beiden Stimmen deutlich. Dies spiegelt sich in der Prominenz des cepstralen Gipfels wider: Bei modaler Stimme sticht der cepstrale Gipfel deutlich hervor, während er sich bei heiserer Stimme kaum von den benachbarten Queffrenzen abhebt.

Die Prominenz des cepstralen Gipfels wird gemessen als Differenz zwischen der Amplitude des Gipfels und der Amplitude des korrespondierenden Punktes (gleiche Queffrenz) auf der Regressionslinie des Cepstrums; dies wird in Abbildung 14 illustriert. Dargestellt sind die Cepstren gehaltener Vokale mit modaler Stimme (oben), mit mittelgradig heiserer Stimme (mitte) sowie mit hochgradig heiserer Stimme (unten). Die Prominenz des cepstralen Gipfels (*Cepstral Peak Prominence/CPP*) ist jeweils links des Graphen abgetragen als Differenz zwischen der Gipfelamplitude und der Amplitude des Punktes auf der Regressionslinie mit der selben Queffrenz (*peak queffrenzy*). Wird die gleiche Messung im geplätteten Cepstrum und dessen Regressionslinie durchgeführt (jeweils rechts des Graphen abgetragen), erhält man die Prominenz des geglätteten cepstralen Gipfels (*Smoothed Cepstral Peak Prominence/CPPS*).

Die Interpretation des CPPS-Wertes ist einfach: Je höher der Wert desto besser. Gesunde Stimmen erreichen i.d.R. CPPS-Werte oberhalb von 14,47 dB [18].

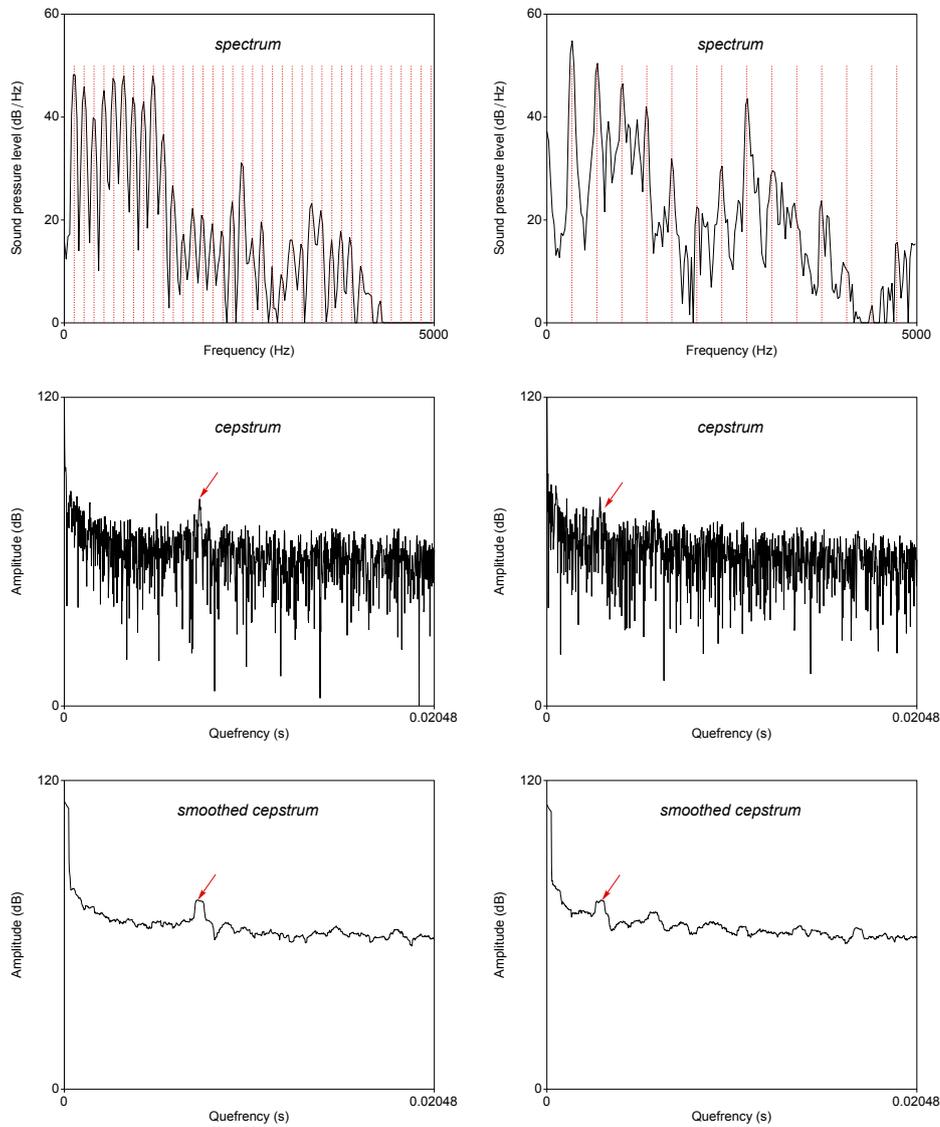


Abbildung 13: Spektrum (oben), Cepstrum (mitte) und geglättetes Cepstrum (unten) eines angehaltenen Vokals; modale Stimme (links) und heisere Stimme (rechts); der cepstrale Gipfel ist jeweils durch rote Pfeile gekennzeichnet; harmonische Gitterlinien im Spektrum (rot) helfen bei der Beurteilung der Periodizität.

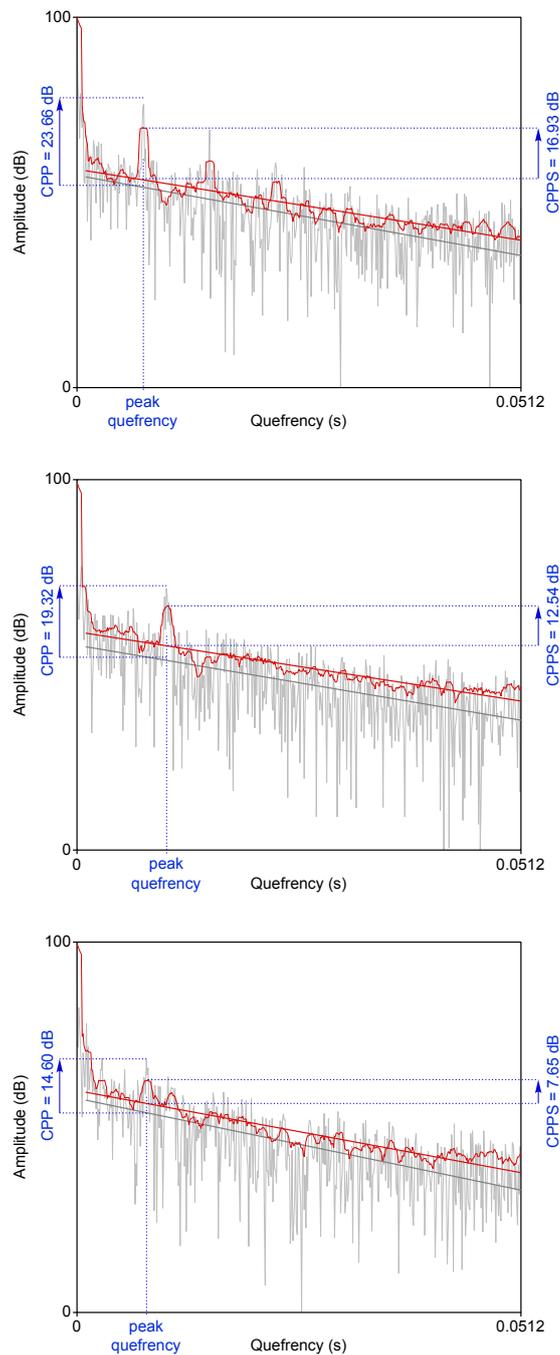


Abbildung 14: Messung der (Smoothed) Cepstral Peak Prominence (modale Stimme (oben), mittelgradig heiser (mitte), hochgradig heiser (unten)); dargestellt ist jeweils das ungeglättete Cepstrum mit Regressionslinie (grau) sowie das geglättete Cepstrum mit Regressionslinie (rot); die Cepstral Peak Prominence (CPP) ist links abgetragen, die Smoothed Cepstral Peak Prominence (CPPS) rechts.

6.8 Grundfrequenz

[Pitch mean/min/max/sd, Range]

Für cs- und sv-Stimmproben (nicht jedoch für die kombinierte mx-Stimmprobe) berechnet VOXplot auch fünf einfache Grundfrequenz-Parameter: Mittelwert (*Pitch mean*), Minimum und Maximum (*Pitch min* und *Pitch max*), die Standardabweichung der Grundfrequenz (*Pitch sd*) sowie den Stimmumfang in Halbtonschritten (*Range*). Diese Parameter erscheinen nur in der textuellen Ergebnisausgabe (s. Abschnitt 4.3.1) und im CSV-Export (s. Abschnitt 4.3.3, nicht jedoch im graphischen Stimmprofil).

Die Bestimmung der Grundfrequenz aus dem akustischen Signal ist keineswegs trivial und es gibt unzählige Algorithmen mit spezifischen Vor- und Nachteilen, die sich dieser Aufgabe widmen [46]. VOXplot verwendet einen Praat-Algorithmus, der von Paul Boersma entwickelt wurde und in [23] beschrieben ist.

Grundsätzlich gilt: (1) **Unterschiedliche Grundfrequenzalgorithmen führen zu unterschiedlichen Ergebnissen.** In einem gewissen Rahmen müssen diese Unterschiede hingenommen werden. Anders formuliert: Algorithmisch ermittelten Grundfrequenzwerten ist eine gewisse Unschärfe zuzugestehen, was man beim Umgang mit solchen Daten immer im Hinterkopf behalten sollte. (2) **Grundfrequenzalgorithmen machen Fehler.** Daher erfordert die Interpretation algorithmisch ermittelter Grundfrequenzwerte Erfahrung und die Ergebnisse müssen stets auf Plausibilität geprüft werden.

Unter anderem die folgenden Gegebenheiten sollten bei einer Plausibilitätsprüfung berücksichtigt werden:

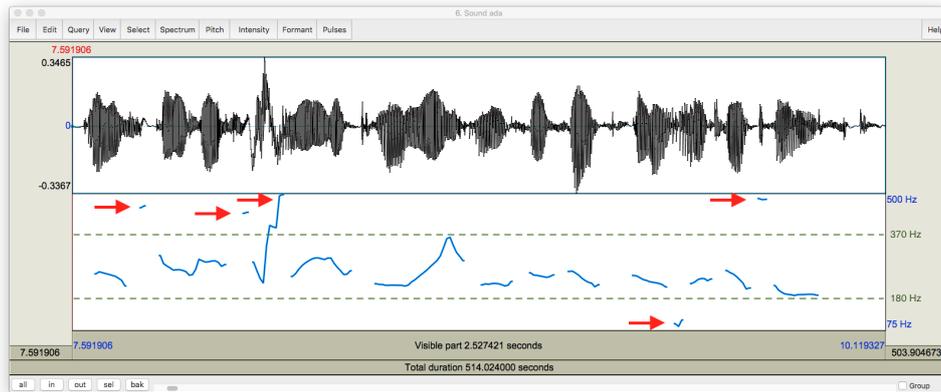
Der Höreindruck: Stimmt der Höreindruck (steigender/fallender/gleichbleibender Stimmton) mit den ermittelten Werten überein?

Segmentale Struktur: Bestimmte Segmente bzw. Segmentübergänge stellen Grundfrequenzalgorithmen regelmäßig vor Probleme: (insbesondere stimmhafte) Plosive, stimmhafte Frikative, Vibranten, das Aufeinandertreffen von Sonorant und Vokal an Wortgrenzen (Laryngalisierung) usw.

Phrasale Struktur: Bei tief endenden Phrasen neigen viele Sprecher und Sprecherinnen zu Laryngalisierung; dies führt oft zu unrealistisch tiefen F_0 -Werten.

Physiologie: Sprecher/Sprecherinnen nutzen typischerweise eine bestimmte Sprechstimmlage. Diese stellt einen Ausschnitt dar aus dem gesamten Stimmtonbereich, den ein Sprecher/eine Sprecherin aufgrund der physiologischen Gegebenheiten in der Lage ist zu produzieren. Moderate Abweichungen von der typischen Sprechstimmlage sind somit möglich, aber unüblich und daher zunächst einmal 'verdächtig'. Extreme Abweichungen sind mehr als verdächtig und können/müssen ignoriert werden.

Standardmäßig sucht Praat im Bereich zwischen 75 Hz und 500 Hz nach Grundfrequenzwerten. Damit werden zwar einerseits fast alle erwachsenen Sprecher und Sprecherinnen abgedeckt, andererseits ist der Bereich für die meisten Stimmen viel zu breit, d.h. der Algorithmus hat zuviele Freiheitsgrade und wird geradezu dazu eingeladen, fehlerhafte Extremwerte zu finden. Hier ein Beispiel:



Die zwei grünen gestrichelten Linien markieren den visuell recht eindeutig erkennbaren tatsächlichen Stimmumfang, die roten Pfeile markieren fehlerhafte Extremwerte. Ohne weiteren Eingriff liefert der Praat-Algorithmus in diesem Beispiel unrealistische Werte, zumindest für das Minimum, das Maximum und den Stimmumfang; der Mittelwert ist i.d.R. unproblematisch. Um dies zu vermeiden, implementiert VOXplot ein Verfahren von Daniel Hirst [49], um automatisch einen angepassten Analysebereich zu finden und verlässlichere Ergebnisse zu liefern. Die Grundfrequenzparameter (Mittelwert, Minimum, Maximum, Standardabweichung und Stimmumfang) werden in der Ergebnisausgabe gelistet (s. Abschnitt 4.3.1), nicht jedoch im Stimmprofil.

Nun zu den Parametern und ihrer Interpretation. Der **Mittelwert** erlaubt Rückschlüsse auf die mittlere Sprechstimmlage eines Sprechers oder einer Sprecherin. Soll es primär um die Bestimmung der mittleren Sprechstimmlage gehen, eignet sich hier sicher eine längere Äußerung besser als ein gehaltener Vokal.

Die **Standardabweichung** ist ein Maß dafür, wie sehr die Grundfrequenz im Verlauf der analysierten Äußerung von der mittleren Grundfrequenz abweicht. Bei einem gehaltenen Vokal ist hier sicherlich ein kleinerer Wert besser (d.h. die Tonhöhe wird gehalten), wird jedoch eine längere Äußerung analysiert (z.B. *Nordwind und Sonne*), ist ein kleinerer Wert eher ein Indiz für eine eingeschränkte Tonhöhenmodulation, also monotones Sprechen.

Aus dem **Minimum** und **Maximum** lässt sich der von der Sprecherin oder dem Sprecher genutzte **Stimmumfang** ableiten. Auch hier gilt, dass bei einem gehaltenen Vokal eher eine geringe Variationsbreite positiv zu bewerten ist (Tonhaltefähigkeit), bei einer längeren Äußerung dagegen eine große Variationsbreite ('lebendige' Stimme). VOXplot zeigt den Stimmumfang nicht in Hertz sondern in Halbtonschritten (st); die Umrechnung erfolgt nach der folgenden Formel: $st = \frac{12}{\ln(2)} \ln\left(\frac{f_{max}}{f_{min}}\right)$

Bei einem vorgelesenen Text sollten Sprecher/Sprecherinnen mit gesunder Stimme zumindest eine knappe Oktave (1 Oktave = 12 Halbtonschritte) erreichen.

Im Einzelfall kann die Bewertung der Grundfrequenzparameter natürlich von der hier vorgeschlagenen Einschätzung abweichen, abhängig davon, welches Ziel beispielsweise in einer Stimmtherapie verfolgt wird.

7 Multiparametrische Stimmqualitätsindizes

Aus den vorausgegangenen Abschnitten ist vermutlich deutlich geworden: Die akustische Analyse der Stimmqualität ist nicht trivial. Es gibt nicht *den* einen Parameter, der klar und eindeutig mit auditiven Stimmqualitätsurteilen korrelieren und sich als objektives numerisches Maß eignen würde. Daraus folgt, dass für eine akustische Stimmanalyse immer mehrere Parameter zu berücksichtigen und zueinander in Beziehung zu setzen sind. Dies versuchen multiparametrische Indizes zu leisten, die in aufwändigen statistischen Analysen relevante akustische Parameter und deren Gewichtung ermitteln und zu einem Index kombinieren.

VOXplot integriert aktuell zwei Stimmqualitätsindizes, die in den letzten Jahren entwickelt wurden und deren Aussagekraft in zahlreichen Studien nachgewiesen werden konnte: Den *Acoustic Voice Quality Index* (AVQI), der mit dem Gesamteindruck der Heiserkeit einer Stimme korreliert, und den *Acoustic Breathiness Index* (ABI) als spezifisches Maß für den Grad der Behauchtheit einer Stimme. Ein Index, der die Dimension Rauigkeit abdeckt, existiert bisher leider nicht [10].

7.1 Acoustic Voice Quality Index

[AVQI]

Der Acoustic Voice Quality Index ist ein multiparametrischer, akustisch basierter Index für die objektive Bewertung von Heiserkeit. Die Signalverarbeitung, Analyse und Auswertung wird in seiner Ursprungsform mit Hilfe der Software Praat und einem dazu individuell entwickelten Skript ermöglicht [8]. Der AVQI umfasst die folgenden sechs akustischen Parameter: Harmonics-to-Noise Ratio, Shimmer in %, Shimmer in dB, allgemeine Steigung des Long-Term Average Spektrum (Slope), Steigung der Regressionslinie durch das Long-Term Average Spektrum (Tilt) und Smoothed Cepstral Peak Prominence (CPPS; CPPS leistet den Hauptbeitrag in diesem Modell) [68].

Für die AVQI-Analyse werden drei Sekunden eines angehaltenen Vokals [a:] ohne Stimman-satz und -absatz und eine Sprechprobe von fortlaufender Sprache (in der Regel eine Passage eines phonetisch ausbalancierten Textes) verwendet und in einer kombinierten Form analysiert. Beide Stimmproben fließen in das endgültige AVQI-Ergebnis ein, da die Bewertung des angehaltenen Vokals alleine keine ausreichende ökologische Validität besitzt [69]. Seit seiner Entwicklung im Jahr 2010 hat der AVQI zwei wichtige und validierte Anpassungen erfahren [21, 52]. Erstens wurde die Struktur der Signalverarbeitungsanalyse vollständig in Praat zusammengefasst (AVQI v.02), statt wie bisher mit zwei Software-Programmen. Zweitens wurde die Dauer des fortlaufenden Sprachanteils sprachspezifisch auf die Silbenzahl standardisiert, die etwa drei Sekunden an stimmhaften Segmenten ergibt (AVQI v.03). Die beiden Stimmproben-Typen werden somit gleich gewichtet, so dass für das endgültige AVQI-Ergebnis eine ausgewogenere Beurteilung stattfinden kann. VOXplot implementiert mit AVQI v.03 die aktuellste AVQI-Version, die neben den oben genannten Vorteilen auch hinsichtlich der diagnostischen Validität signifikant verbessert ist gegenüber früheren AVQI-Versionen [52].

Das Ergebnis der AVQI-Analyse ist ein Wert i.d.R. zwischen 0 und 10, wobei ein höherer Wert eine stärkere Heiserkeit anzeigt und vice versa (in Ausnahmefällen können die Grenzwerte 0 und 10 unter- oder überschritten werden). Der AVQI-Cutoff-Score, d.h. der Wert, der zwischen normalen und heiseren Stimmen unterscheidet, variiert zwischen den verschiedenen AVQI-Versionen [21, 92] und zwischen verschiedenen Sprachen bzw. kulturellen Hintergründen [21, 37]. Aktuell liegen sprachspezifische Validierungsstudien für den AVQI v.03 für 12 Sprachen vor – darunter Deutsch [1, 9, 15, 29, 32, 39, 40, 51, 55, 57, 58, 84, 94]. Diese 12 Sprachen können in den **VOXplot**-Einstellungen als Analysesprache ausgewählt werden (vgl. Abschnitt 5.3).

Der AVQI hat einen hohen klinischen Nutzen und Meta-Analysen zeigen gute Ergebnisse hinsichtlich der diagnostischen Genauigkeit, der Kriteriumsvalidität zur perceptiven Beurteilung und der Fähigkeit, Stimmveränderungen im Zeitverlauf nach verschiedenen Stimmbehandlungen zu erkennen [21, 52]. Die diagnostische Genauigkeit des AVQI ($n = 11696$), die Kriteriumsvalidität ($n = 10683$) und die Messungen bei Stimmveränderungen ($n = 734$) wurden mit unterschiedlich vielen Stimmproben ausgewertet – in einer der bislang größten Studien im Bereich der Stimmqualitätsanalysen [21]. Die Ergebnisse dieser Meta-Analyse zur Validität des AVQI sind die Grundlage zur Einschätzung der diagnostischer Qualität des AVQI, die im weiteren beschrieben wird.

Die Fähigkeit des AVQI zur Unterscheidung zwischen normalen und heiseren Stimmen zeigte eine Übereinstimmung von zusammengefasst 93,7%, ermittelt durch eine Grenzwertoptimierungskurvenstatistik. Es konnten ausreichend hohe Werte der gepoolten Sensitivität von 83% (95% Konfidenzintervall: 82% - 83%) und der gepoolten Spezifität von 89% (95% KI: 88% - 90%) ermittelt werden, sowie eine positive Likelihood-Ratio von 7,75 (95% KI: 6,04 - 9,95) und eine negative Likelihood-Ratio von 0,20 (95% KI: 0,17 - 0,23). Außerdem wurden für den AVQI hohe Korrelationen mit einem gewichteten mittleren Korrelationskoeffizienten von $r = 0,796$ für die Validität der Veränderungssensitivität bestimmt.

Einen allgemeinen Cutoff-Score aus den Daten dieser Meta-Analyse zu erschließen ist nicht möglich, da die gepoolte Sensitivität und Spezifität aus allen Versionen und Entwicklungsschritten des AVQI zusammengefasst worden sind. Daher verwendet **VOXplot** die in den jeweiligen Validierungsstudien ermittelten sprachspezifischen Cutoff-Scores. Der für das Deutsche mit AVQI v.03 ermittelte Cutoff-Score liegt bei 1,85 [15], d.h. ein AVQI-Wert über 1,85 indiziert eine heisere Stimme, wobei der AVQI-Wert mit dem Grad der Heiserkeit ansteigt.

Weitere Aspekte, die in Studien zum AVQI untersucht worden sind:

- ▷ Das Geschlecht hat keinen Einfluss auf den AVQI-Wert, das Alter kann einen minimalen Einfluss haben [19, 52, 54].
- ▷ Es wurden akzeptable Werte für den AVQI bei der Beurteilung der Stimme im Rahmen der Erkennung von Stimmstörungen (normal versus organisch/funktionelle Stimmstörungen) ermittelt [20] und auch eine mäßige bis hohe Korrelation mit standardisierten Fragebögen zur Einschätzung von Beeinträchtigungen der Stimme durch Stimmprobleme festgestellt [84, 95].

- ▷ Der AVQI zeigt sich als robust gegenüber Schwankungen in der Raumakustik, den Hintergrundgeräuschen und der Mikrofonqualität [26, 95]. Dennoch sollten diese Faktoren kontrolliert werden (vgl. Kapitel 8).
- ▷ Der AVQI wurde zwar entwickelt, um Heiserkeit abzubilden, es hat sich jedoch gezeigt, dass Anomalien der Stimmqualität in der Dimension Behauchtheit stärker erfasst werden als Anomalien in der Dimension der Rauigkeit [11]; hier besteht weiterer Forschungsbedarf.
- ▷ Da **VOXplot** eine integrierte, validierte Praat-Version enthält und nötige Änderungen – falls notwendig – vor der Veröffentlichung ausgiebig getestet werden, sind die in [92] formulierten Bedenken hinsichtlich Praat-Updates, die einen Einfluss auf die Analysefunktionen von AVQI haben können, unbegründet.

7.2 Acoustic Breathiness Index

[ABI]

Der Acoustic Breathiness Index (ABI) analysiert den Behauchtheitsgrad einer Stimme und ist ebenfalls ein multiparametrischer akustischer Index. Der ABI verwendet das gleiche statistische Verfahren für die Entwicklung des Index und das gleiche Sprachmaterial für die Analyse (3 Sekunden eines gehaltenen Vokals [a:] ohne Stimmansatz und -absatz und eine (sprachspezifische) Sprechprobe von fortlaufender Sprache) wie der AVQI v.03. Auch das Beurteilungsschema auf einer Skala von 0-10 (je höher der ABI Wert, desto behauchter die Stimme) ist vergleichbar. Im Gegensatz zum AVQI bildet der ABI den Grad der Behauchtheit einer Stimme ab. Zur Berechnung des Index werden neun akustische Parameter herangezogen: CPPS, Jitter %, GNE, Hf Noise, HNR-D, H1-H2, Shimmer %, Shimmer dB und PSD [17].

Eine Meta-Analyse bestätigt auch dem ABI einen hohen klinischen Nutzen und gute diagnostische Genauigkeit [13]. Die Analyse hat 3603 Stimmproben berücksichtigt, davon 467 stimmgesunde und 3136 stimmgestörte Stimmproben. Die Ergebnisse dieser Meta-Analyse dienen als Grundlage für die Einschätzung der diagnostischen Qualität des ABI, die im weiteren beschrieben wird.

Die gepoolte Sensitivität liegt bei 84% (95% KI: 83% - 85%), die gepoolte Spezifität bei 92% (95% KI: 89% - 94%). Die Fähigkeit zur Unterscheidung zwischen der Abwesenheit und dem Vorhandensein von Behauchtheit durch ABI zeigte eine Übereinstimmung von zusammengefasst 94% ermittelt durch eine Grenzwertoptimierungskurvenstatistik. Da in allen inkludierten Studien der Meta-Analyse vergleichbare Standards gewährleistet waren, konnte ein allgemeiner, gewichteter ABI-Cutoff-Score von 3,40 aus sechs Sprachen ermittelt werden (Sensitivität: 86%, 95% KI: 84% - 87%; Spezifität: 90%, 95% KI: 88% - 92%). Dennoch verwendet **VOXplot** wie beim AVQI auch für den ABI die in sprachspezifischen Validierungsstudien ermittelten Cutoff-Scores (Deutsch: 3,42); aktuell liegen Validierungsstudien für 11 Sprachen vor [1, 14, 15, 17, 29, 32, 38, 50, 53, 56, 58].

Die Kriteriumsvalidität zwischen ABI und einer auditiv-perzeptiven Beurteilung der Behauchtheit reicht von $r = 0,746$ bis $r = 0,890$ [13] und die Fähigkeit des ABI, therapiebedingte Veränderungen der Stimmqualität nachzuweisen, wird mit Korrelationswerten

von $r = 0,71$ bis $r = 0,88$ als ausreichend angesehen [14, 50]. Alter, Geschlecht und Grad der Rauigkeit haben keinen signifikanten Einfluss auf die ABI-Analyse, weder auf Basis von natürlichen [50] noch auf Basis von synthetischen Stimmen [11].

8 Recording: Tipps und Empfehlungen

Zahlreiche Faktoren haben Einfluss auf die Qualität und die Vergleichbarkeit einer akustischen Stimmanalyse. Viele dieser Faktoren haben mit dem Aufnahme-Setting zu tun, also mit den Umständen, unter denen eine Stimmprobe, die einer akustischen Analyse unterzogen werden soll, aufgenommen wird. Dazu zählen Hardware-Aspekte (Mikrofon, Mikrofon-Verstärker, AD-Wandler), aber auch z.B. die Raumakustik, das Hintergrundrauschen usw. Im Folgenden werden diese Punkte kurz angesprochen, kombiniert mit einigen Tipps und Empfehlungen (vgl. [5, 26]).

Pragmatisch betrachtet kann man m.E. zwei Qualitätsstufen unterscheiden: (1) Ausreichende Qualität, um Vergleichbarkeit und Zuverlässigkeit der *eigenen* Messungen zu gewährleisten und (2) höchste Qualität, um Vergleichbarkeit mit Studienergebnissen zu erreichen; Stufe (1) ist naturgemäß mit geringerem Aufwand und billiger zu erreichen als Stufe (2). Unabhängig davon, in welche Qualitätsstufe Sie investieren wollen, ist die wichtigste Grundregel für Vergleichbarkeit: **Variieren Sie das Aufnahme-Setting möglichst wenig bis gar nicht**. Nachdem Sie also einmal ein für Sie optimales Setting erarbeitet haben, dokumentieren Sie dieses am besten und bleiben Sie dann dabei, denn nur dann sind zuverlässige Vergleiche zwischen unterschiedlichen Messzeitpunkten möglich.

8.1 Geräte

Mikrofon

Bauart: In Frage kommen Stand- und Kopfbügelmikrophone. Standmikrophone sind etwas anfälliger für Umgebungslärm und der gleichbleibende Abstand zur Schallquelle (≈ 30 cm) ist schwerer zu kontrollieren. Kopfbügelmikrophone sind weniger anfällig für Umgebungslärm (Empfehlung: weniger als 50 dBA) und der gleichbleibende Abstand zur Schallquelle ist gewährleistet; der nachteilige Nahbesprechungseffekt wird von hochwertigen Kopfbügelmikrofonen i.d.R. kompensiert.

Funktionsprinzip: Kondensatormikrophone sind dynamischen Mikrofonen eindeutig vorzuziehen (besseres Impulsverhalten, weiterer Übertragungsbereich usw.); idealerweise ein Kondensatormikrofon mit einer Empfindlichkeit > 60 dB.

Charakteristik: Nierencharakteristik, wegen der besseren Fokussierung auf die Schallquelle (anstatt auf den Umgebungslärm).

Frequenzbereich: Idealerweise wird der gesamte Bereich der menschlichen Stimme abgedeckt: 20 Hz – 20 kHz.

Frequenzverlauf: Möglichst flach, d.h. höchstens 2 dB Abweichung zwischen 20 Hz und 8 kHz, idealerweise bis 20 kHz.

Äquivalenzschalldruckpegel: (Eigenrauschen des Mikrofon) max. 25 dB(A)

Maximaler Schalldruckpegel: min. 126 dB(A)

Als Beispiel seien zwei Mikrophone genannt, die diese Anforderungen erfüllen und mit Preisen um 120 € ein gutes Preis-Leistungsverhältnis haben: Das Standmikrofon AKG P120 (getestet in [26]) und das Kopfbügelmikrofon AKG C 544 L (verwendet u.a. in [15, 57]). Mit diesen beiden Mikrofonen kann man also Qualitätsstufe (2) erreichen (allerdings muss man in weiteres Zubehör investieren, s.u.). Natürlich gibt es auch von anderen Herstellern sehr gute Mikrophone im Preisbereich zwischen 100 € und 200 € (Audio-Technica, Behringer, Presonus, Røde etc.).

Bei günstigeren Mikrofonen (Stufe (1)) muss man v.a. Abstriche beim Frequenzbereich und Frequenzverlauf hinnehmen und man verliert weitgehend die Vergleichbarkeit mit Studienergebnissen und den dort u.U. postulierten Norm- bzw. Grenzwerten. Bottalico *et al.* zeigen dies am Beispiel eines sehr günstigen Logitech Headsets [26]. Die gleiche Studie kommt zu dem Ergebnis, dass Jitter und CPPS am wenigsten Abhängigkeit vom verwendeten Mikrofon zeigen, gefolgt von Shimmer und HNR. Die größte Abhängigkeit vom verwendeten Mikrofon zeigt der Parameter Spektrale Neigung und der multiparametrische AVQI.

Einfache externe Mikrophone, die per Klinkestecker mit dem Computer verbunden werden, sollte man generell meiden. Die Mikrophone sind sehr weit von den oben angegebenen Spezifikationen entfernt und auch der (zunächst einmal analoge) Signalweg im Computer erreicht oft nicht die gewünschte Qualität. Idealerweise bekommt es der Computer nur mit bereits digitalisierten Signalen zu tun, denen er nichts mehr antun kann: Digitale Signale behalten ihre Ursprungsqualität im Prinzip bis in alle Ewigkeit... Darum sollte man bei der Aufnahme und der Digitalisierung möglichst sorgfältig vorgehen; danach kann nicht mehr viel passieren.

Mikrofonverstärker / AD-Wandler

Kondensatormikrophone müssen mit einer Gleichspannung gespeist werden (sog. *Phantomspannung* bzw. *Phantomspannung*) und das analoge Mikrophonsignal muss verstärkt und digitalisiert werden (Analog-Digital-Wandlung/AD-Wandlung). Sogenannte USB-Mikrophone haben all diese Funktionen eingebaut, sie benötigen keine weiteren Komponenten. Sie werden via USB direkt mit dem Computer verbunden und liefern ein fertig digitalisiertes Signal. Die Aufnahme wird direkt am Computer mit Praat oder einem beliebigen anderen Audio-Editor gesteuert und als Audio-Datei abgespeichert. USB-Mikrophone gibt es als einfache und günstige Varianten wie z.B. das oben erwähnte Logitech-Headset, es gibt aber auch sehr ordentliche Großmembran-Kondensator-Modelle, die oft unter der Bezeichnung *Podcast-Mikrofon* vermarktet werden (ab 50 €). Bei günstigen Modellen wird sowohl an der Übertragungsqualität des Mikrophons als auch an der Qualität des Mikrofonverstärkers und des AD-Wandlers gespart. Nachteile dieser Bauform: Kein Austausch/Upgrade einzelner Komponenten. Außer [26], die das Logitech-Headset (eher als abschreckendes Beispiel) in ihren Test aufgenommen haben, ist mir keine Stimm-Studie bekannt, die ein USB-Mikrofon verwendet hätte, d.h. die Vergleichbarkeit ist eingeschränkt. Generell gilt: Bevor ein Mikrofon für die akustische Stimmanalyse verwendet wird, sollte anhand der technischen Daten, die in der Bedienungsanleitung zu finden sind, sichergestellt werden, dass die technischen Anforderungen (vgl. Seite 53) möglichst umfassend erfüllt sind.

Hochwertige Mikrophone wie z.B. die erwähnten AKG-Modelle sind erstmal nur Mikrophon; sie benötigen weitere Komponenten zum Funktionieren. Im Prinzip kann man alle notwendigen Komponenten (Phantomspeisung, Mikrophon-Verstärker, AD-Wandler) einzeln und jeweils in beliebiger Qualität (und fast beliebig hohen Preisen) kaufen und hintereinander stöpseln. Weniger aufwändig, günstiger und üblich sind sog. USB-Interfaces. Das sind kleine Kästchen, die die geforderten Funktionen vereinen, also das per XLR-Stecker angeschlossene Mikrophon mit Spannung versorgen sowie das analoge Signal verstärken und digitalisieren, und das digitalisierte Signal dann per USB an den Computer weiterreichen. Schon die günstigsten Geräte (ab rund 100 €) der Markenhersteller (Audient, Focusrite, Native Instruments, Steinberg usw.) liefern sehr gute Qualität. Ein Test solcher Interfaces mit speziellem Fokus auf akustischer Stimmanalyse, wie es ihn exemplarisch für Mikrophone gibt ([26]), ist mir leider nicht bekannt.

Falls Qualitätsstufe (1) ausreicht, kommt evtl. auch ein mobiler digitaler Recorder in Frage (von Zoom, Tascam usw.). Damit sind gute Aufnahmen unabhängig vom Computer möglich. Die Mini-Recorder werden mit Batterie oder Akku betrieben und vereinen ein ordentliches Kondensator-Mikrophon, Verstärker, AD-Wandler und Speichermedium in einem Gerät und sind somit recht universell verwendbar. Kosten: ab ca. 90 € aufwärts.

Unabhängig von der verwendeten Hard- und Software ist bei der AD-Wandlung auf die folgenden Einstellungen zu achten:

- ▷ Abtastrate: üblich sind 44,1 kHz
- ▷ Auflösung: mindestens 16 bit
- ▷ Speicherformat: ein unkomprimiertes PCM-Format, z.B. das wav-Format (keine Formate verwenden, die verlustbehaftet komprimieren, wie z.B. MP3!)

8.2 Raumakustik

Umgebungsärm reduzieren

Störgeräusche jeglicher Art müssen während der Aufnahme vermieden werden. Quellen für Störgeräusche sind z.B. Computerlüfter, Netzteile (z.B. von Schreibtischlampen oder Bildschirmen), Papier (Aufnahme von abgelesenem Material), Trittschall, Jalousien, Faxgeräte usw. Solche Störquellen sollten entweder ganz vermieden werden (z.B. Lesetexte auf einem Notenständer platzieren) oder zumindest abgeschirmt werden (Styroporplatten, Kissen etc.) oder ihr Einfluss sollte durch möglichst große Distanz zum Mikrophon minimiert werden. Empfehlung: Die Differenz zwischen dem Umgebungsärmpegel und dem Pegel des Stimmsignals (Signal-Rausch-Verhältnis) sollte min. 30 dB(A) betragen, besser mehr. Das Signal-Rausch-Verhältnis eines konkreten Aufnahme-Settings kann mithilfe von **VOXplot** einfach evaluiert werden; mehr dazu in Abschnitt 4.2.1.

Nachhall reduzieren

Hall sollte ebenfalls vermieden werden, z.B. indem man Sprecher*innen nicht direkt vor einer Wand oder einem Bildschirm platziert, indem man für die Aufnahme einen Raum mit Teppichboden bevorzugt, indem man einen Aufnahmebereich mit möglichst wenig Glasflächen und möglichst vielen 'weichen' Möbeln (Oberflächenmaterialien mit hoher Schallabsorption) wählt usw.

8.3 Stimmverhalten

Zu guter Letzt sollte bei der Aufnahme von Stimmproben darauf geachtet werden, dass die Lautstärke bei der gehaltenen Phonation eines Vokals vergleichbar zur habituellen Sprechlautstärke der untersuchten Person ist. Die habituelle Sprechlautstärke lässt sich am besten bei fortlaufendem Sprechen (z.B. *Nordwind und Sonne*) bestimmen. Wird der gehaltene Vokal deutlich lauter phoniert, kann sich das Ergebnis der akustischen (und auch der auditiven) Analyse u.U. drastisch verändern [60, 98].

8.4 Mund-Nasen-Masken

Eine Meta-Analyse auf der Basis von 9 Studien (mit insgesamt 422 Versuchspersonen mit und ohne Stimmstörung) kommt zu dem Ergebnis, dass das Tragen einer Mund-Nasen-Maske durch die untersuchte Person keinen signifikanten Einfluss auf die Messergebnisse einer akustischen Stimmanalyse hat [12]. Die Meta-Analyse hat 6 akustische Parameter (Jitter, Shimmer, HNR, CPPS, AVQI und die habituelle Sprechlautstärke (SPL)) und 2 Maskentypen berücksichtigt (medizinische Gesichtsmaske/"OP-Maske" und FFP2-Maske). Allerdings ergab der indirekte Vergleich zwischen den beiden Maskentypen bei 4 von 6 Parametern signifikante Unterschiede (Jitter, HNR, CPPS und SPL). Sofern also das Tragen einer Maske erforderlich oder erwünscht ist, können solche Aufnahmen anscheinend ohne weiteres verwendet werden, es sollte jedoch darauf geachtet werden, dass stets der gleiche Maskentyp benutzt wird.

Literatur

- [1] M. Aghajanzadeh, S. Saeedi, S. Jalaie, K. Esarian und 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 und N. Amir. "A Clinical Comparison between Two Acoustic Analysis Softwares: MDVP and Praat". In: *Biomedical Signal Processing and Control* 4.3 (2009), S. 202–205.
- [3] S. N. Awan, A. Giovinco und 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 und 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), S. 430–439.
- [5] B. Barsties und M. De Bodt. "Assessment of Voice Quality: Current State-of-the-Art". In: *Auris, Nasus, Larynx* 42.3 (2015), S. 183–188.
- [6] B. Barsties, U. Hoffmann und Y. Maryn. "Spektrografische Stimmtypenklassifizierung Zur Beurteilung Der Stimmqualität". In: *Laryngo-Rhino-Otol* 95 (2016), S. 105–111.
- [7] B. Barsties und Y. Maryn. "Der Acoustic Voice Quality Index in Deutsch. Ein Messverfahren Zur Allgemeinen Stimmqualität". In: *HNO* 60.8 (2012), S. 715–720.
- [8] B. Barsties und Y. Maryn. "External Validation of the Acoustic Voice Quality Index Version 03.01 With Extended Representativity". In: *Annals of Otolaryngology & Laryngology* 125.7 (2016), S. 571–583.
- [9] B. Barsties und Y. Maryn. "The Improvement of Internal Consistency of the Acoustic Voice Quality Index". In: *American Journal of Otolaryngology* 36.5 (2015), S. 647–656.
- [10] B. Barsties v. Latoszek, M. De Bodt, E. Gerrits und Y. Maryn. "The Exploration of an Objective Model for Roughness with Several Acoustic Markers". In: *Journal of Voice* 32.2 (2018), S. 149–161.
- [11] B. Barsties v. Latoszek, M. Englert, J. C. Lucero und 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 und 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 und 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), S. 31–40.
- [14] B. Barsties v. Latoszek und B. Lehnert. "Interne Validität des Acoustic Voice Quality Index Version 03.01 und des Acoustic Breathiness Index". In: *Laryngo-Rhino-Otologie* 97.09 (2018), S. 630–635.

- [15] B. Barsties v. Latoszek, B. Lehnert und 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 und M. De Bodt. "A Meta-Analysis: Acoustic Measurement of Roughness and Breathiness". In: *Journal of Speech, Language, and Hearing Research* (2018), S. 1–26.
- [17] B. Barsties v. Latoszek, Y. Maryn, E. Gerrits und 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 und B. Lehnert. "Advances in Clinical Voice Quality Analysis with VOXplot". In: *Journal of Clinical Medicine* 12.4644 (2023).
- [19] B. Barsties v. Latoszek, N. Ulozaite-Staniene, Y. Maryn, T. Petrauskas und 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), S. 340–345.
- [20] B. Barsties v. Latoszek, N. Ulozaite-Staniene, T. Petrauskas, V. Uloza und Y. Maryn. "Diagnostic Accuracy of Dysphonia Classification of DSI and AVQI". In: *The Laryngoscope* 129 (2019), S. 692–698.
- [21] C. Batthyany, B. Barsties v. Latoszek und 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*. Hrsg. von M. K. Huffman und R. A. Krakow. Academic Press, 1993, S. 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*. Bd. 17. 1193. Amsterdam. 1993, S. 97–110.
- [24] P. Boersma. "Should Jitter Be Measured by Peak Picking or by Waveform Matching?" In: *Folia Phoniatica et Logopaedica* 61.5 (2009), S. 305–308.
- [25] B. P. Bogert, M. J. R. Healy und J. W. Tukey. "The Quefrency Alanysis of Time Series for Echoes: Cepstrum, Pseudo-Autocovariance, Cross-Cepstrum, and Saphe Cracking". In: *Proceedings of the Symposium on Time Series Analysis*. Hrsg. von M. Rosenblatt. Wiley, 1963, S. 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 und A. D. Rubin. "Reproducibility of Voice Parameters: The Effect of Room Acoustics and Microphones". In: *Journal of Voice* 34.3 (2020), S. 320–334.
- [27] L. F. Brinca, A. P. F. Batista, A. I. Tavares, I. C. Gonçalves und 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), S. 282–286.

- [28] M. Brockmann, C. Storck, P. N. Carding und 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), S. 1152–1160.
- [29] A. Castillo-Allendes, J. Codino, L.C. Cantor-Cutiva, C.J. Nudelman, A.D. Rubin, B. Barsties v. Latoszek und 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), S. 343–344.
- [31] P. Dejonckere und 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), S. 988–996.
- [32] J. Delgado Hernandez, N.M. Leon Gomez, A. Jiménez, L.M. Izquierdo und 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), S. 317–326.
- [33] D. D. Deliyski, M Evans und H Shaw. "Influence of Data Acquisition Environment on Accuracy of Acoustic Voice Quality Measurements". In: *Journal of Voice* 19.2 (2005), S. 176–186.
- [34] D. D. Deliyski, H. S. Shaw und M. K. Evans. "Adverse Effects of Environmental Noise on Acoustic Voice Quality Measurements". In: *Journal of Voice* 19.1 (2005), S. 15–28.
- [35] D. D. Deliyski, H. S. Shaw und M. K. Evans. "Influence of Sampling Rate on Accuracy and Reliability of Acoustic Voice Analysis". In: *Logopedics Phoniatrics Vocology* 30.2 (2005), S. 55–62.
- [36] D. D. Deliyski, H. S. Shaw, M. K. Evans und Vesselinov R. "Regression Tree Approach to Studying Factors Influencing Acoustic Voice Analysis". In: *Folia Phoniatria Logop* 58 (2006), S. 274–288.
- [37] M. Englert, B. Barsties v. Latoszek und M Behlau. "The impact of languages and cultural backgrounds on voice quality analyses". In: *Folia Phoniatria et Logopaedica* 74.2 (2022), S. 141–152.
- [38] M. Englert, B. Barsties v. Latoszek, Y. Maryn und M. Behlau. "Validation of the Acoustic Breathiness Index to the Brazilian Portuguese Language". In: *Logopedics Phoniatrics Vocology* 0.0 (2021), S. 1–7.
- [39] M. Englert, B. Barsties v. Latoszek, Y. Maryn und 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 und 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 und J. I. Godino-Llorente. "Cepstral Peak Prominence: A Comprehensive Analysis". In: *Biomedical Signal Processing and Control* 14 (2014), S. 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 und E. Nöth. "Automatic Rating of Hoarseness by Text-Based Cepstral and Prosodic Evaluation". In: (2012), S. 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), S. 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 und J. Hillenbrand. "Quantifying the Cepstral Peak Prominence, a Measure of Dysphonia." In: *Journal of Voice* 28.6 (2014), S. 783–788.
- [46] W. Hess. *Pitch Determination of Speech Signals: Algorithms and Devices*. Bd. 3. Springer, 2012.
- [47] J. Hillenbrand, R. A. Cleveland und R. L. Erickson. "Acoustic Correlates of Breathy Vocal Quality". In: *Journal of Speech, Language, and Hearing Research* 37.4 (1994), S. 769–778.
- [48] J. Hillenbrand und 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), S. 311–321.
- [49] D. Hirst. "The analysis by synthesis of speech melody: from data to models". In: *Journal of speech Sciences* 1.1 (2011), S. 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 und 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), S. 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 und 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 und 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 und 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 und 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 und 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 und 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 und A.-M. Laukkanen. "Akustisen äänenlaatuindeksin (AVQI) version 03.01 validointi suomenkielisille puhujille". In: *Puhe ja kieli* 40.3 (2020), S. 165–182.
- [58] G.-H. Kim, B. Barsties von Latoszek und 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. Syrja, M. Laitala und T. Leino. "Effects of two-month vocal exercising with and without spectral biofeedback on student actors' speaking voice". In: *Logoped Phoniatr Vocol* 29 (2004), S. 66–76.
- [60] E. Lebbe, J. Vestraete, J. DeKlerck und 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 und 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 und R. T. Sataloff. "Reliability of Objective Voice Measures of Normal Speaking Voices." In: *Journal of Voice* 27.2 (2013), S. 170–176.
- [63] S. Y. Lowell, R. H. Colton, R. T. Kelley und 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 und S. A. Mizia. "Predictive Value and Discriminant Capacity of Cepstral- and Spectral-Based Measures during Continuous Speech." In: *Journal of Voice* 27.4 (2013), S. 393–400.
- [65] J. K MacCallum, Y. Zhang und J. Jiang. "Vowel Selection and Its Effects on Perturbation and Nonlinear Dynamic Measures". In: *Folia Phoniatrica et Logopaedica* 63.2 (2011), S. 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 und 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), S. 1609–1619.
- [68] Y. Maryn, M. De Bodt und N. Roy. "The Acoustic Voice Quality Index: Toward Improved Treatment Outcomes Assessment in Voice Disorders". In: *Journal of Communication Disorders* 43.3 (2010), S. 161–174.
- [69] Y. Maryn, P. Corthals, P. Van Cauwenberge, N. Roy und 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), S. 540–555.

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

- [84] T. Pommée, Y. Maryn, C. Finck und 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 und 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 und W. J. Barry. “Instrumental Dimensioning of Normal and Pathological Phonation Using Acoustic Measurements”. In: *Clinical Linguistics & Phonetics* 22.6 (2008), S. 407–420.
- [87] V. Reynolds, A. Buckland, J. Bailey, J. Lipscombe, E. Nathan, S. Vijayasekaran, R. Kelly, Y. Maryn und N. French. “Objective Assessment of Pediatric Voice Disorders with the Acoustic Voice Quality Index.” In: *Journal of Voice* 26.5 (2012), S. 672.
- [88] C. Sauder, M. Bretl und 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), S. 557–566.
- [89] H.-J. Shim, H. R. Jang, H. B. Shin und D.-H. Ko. “Cepstral, Spectral and Time-Based Analysis of Voices of Esophageal Speakers.” In: *Folia phoniatrica et logopaedica* 67.2 (2015), S. 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), S. 477–490.
- [91] A. Sprecher, A. Olszewski, J. J. Jiang und Y. Zhang. “Updating signal typing in voice: addition of type 4 signals”. In: *The Journal of the Acoustical Society of America* 127.6 (2010), S. 3710–3716.
- [92] L. Stappenbeck, Ben Barsties v. Latoszek, B. Janotte und 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), S. 101938.
- [93] J. G. Švec und S. Granqvist. “Guidelines for Selecting Microphones for Human Voice Production Research”. In: *American Journal of Speech-Language Pathology* 19.4 (2010), S. 356–368.
- [94] V. Uloza, T. Petrauskas, E. Padervinskis, N. Ulozaite, B. Barsties und 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 und 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), S. 1737–1745.
- [96] A. P. Vogel und P. Maruff. “Comparison of Voice Acquisition Methodologies in Speech Research”. In: *Behavior Research Methods* 40.4 (2008), S. 982–987.

- [97] A. P. Vogel und 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), S. 431–437.
- [98] I. Yanushevskaya, C. Gobl und A. Ni Chasaide. "Voice quality in affect cueing: does loudness matter?" In: *Frontiers in psychology* 4 (2013).