



Qualitätsbeurteilung von Sprachproben

- 1 [Zielstellung des Forschungsvorhabens](#)
- 2 [Aufnahme von Sprachproben](#)
- 3 [Entwicklung eines Experimentiersystems](#)
 - 3.1 [Hauptprogramm](#)
 - 3.2 [Bestimmung von Frequenzgruppenpegeln](#)
 - 3.2.1 [Festlegung des Analyse-Zeitfensters](#)
 - 3.2.2 [Schnelle FOURIER-Transformation](#)
 - 3.2.3 [Parallelfilterung](#)
 - 3.3 [Berechnung von Erregungspegeln und Kernlautheiten](#)
 - 3.4 [Berechnung der Lautheits-Tonheits-Verteilung](#)
 - 3.5 [Vokalklassifikation](#)
 - 3.6 [Simulation der Empfindungsgröße „Schärfe“](#)
- 4 [Umstellung auf ein Analysesystem](#)
- 5 [Derzeitiger Leistungsstand](#)

Anlage 1: [Tabellen](#)

Anlage 2: [Programmablaufplan](#)

Anlage 3: [Formelzeichenverzeichnis](#)

Anlage 4: [Testsignale](#)

Anlage 5: [Arbeitspläne](#)

[Literaturverzeichnis](#)

Prof. Dr. Ose
Hochschule Ostfalia
Fakultät Elektrotechnik

Forschungsbericht

Thema: Qualitätsbeurteilung von Sprachproben

Bearbeitungszeitraum: Sommersemester 2012 (15. 2. bis 31. 7. 2012)

[Gliederung](#)

1 Zielstellung des Forschungsvorhabens

Im Ergebnis der Bearbeitung eines Forschungsprojektes 1999 bis 2001 wurde ein interaktiver Sprachtrainer für Hör- und Sprechgeschädigte entwickelt [1]. Die Auswertung der Ergebnisse eines Sprechtrainings wurde mit einer einfachen Bewertungsskala vorgenommen. Ein nachfolgendes Projekt 2001 bis 2004 beschäftigte sich mit der PC-gestützten Beurteilung der Qualität von Sprachproben [2], [3].

In beiden Projekten wurden die Sprachproben auf der Grundlage physikalische Signalparameter (Grundfrequenz, Nulldurchgänge, Signalenergie sowie Amplitudenspektrum und Spektrogramm) bewertet. Das humane Perzeptionssystem beurteilt und bewertet Lautsprache aber mit subjektiven Empfindungsgrößen (Lautheits-Tonheits-Verteilung) und Wahrnehmungsgrößen (z.B. Klangattribute; wie Schärfe, Dichte, Volumen, Tonhöhe, Klangschönheit, usw.) [4].

Das Vorhaben zielt darauf ab, ein Programm zur automatischen Bestimmung der Lautheits-Tonheitsverteilung mit LabVIEW zu entwickeln. Mit Kenntnis dieser Verteilung soll versucht werden, ausgewählte Klangattribute auf einer (noch zu definierenden) Skala zu berechnen. Diese Attribute sind dann so miteinander zu verknüpfen, dass sie Informationen über die Qualität einer lautsprachlichen Realisierung tragen. Zur Reduzierung des zu erwartenden Aufwandes wird das Testmaterial auf die deutschen Langvokale begrenzt.

Die Ergebnisse dieses Vorhabens sollen dazu beitragen, neue Erkenntnisse zu Prinzipien der humanen Verarbeitung humaner Lautsprache (automatische Spracherkennung, Sprechererkennung und Qualitätsbeurteilung) zu gewinnen. Im Ergebnis des 2004/2005 eingereichten Forschungsberichtes wurde vom BMBF darauf hingewiesen, dass eine Weiterführung dieser Thematik (Verbesserung / Optimierung der technischen Lösung sowie Erschließung neuer Anwendungsgebiete) erwünscht ist.

2 Aufnahme von Sprachproben

[Gliederung](#)

Zur Bereitstellung des erforderlichen Testmaterials wurde das im Bild 1 dargestellte LabVIEW-Programm entwickelt. Die lautsprachliche Realisierung wird von einem hochwertigen Kondensator-Mikrofon erfasst und über einen Vorverstärker an den PC weitergeleitet. Parallel dazu erfolgt eine messtechnische Bewertung durch einen Schallpegelmess (Spitzenwert). Zur Absicherung einer reproduzierbaren Aufnahme liegt am Arbeitsplatz eine Kalibrierungsvorschrift aus, die vor jeder Aufnahme vom Aufnahmeleiter abgearbeitet ist. Der optimale Abstand des Sprechers zum Mikrofon und die Intensität wird zunächst mit der Aussteuerungsanzeige (Bild 1 – rechts) getestet. Der Grad der Aussteuerung wird als Momentanwert (LED: grün – gelb – rot) und als Spitzenwert über einen roten Balken dargestellt. Die Spitzenwert-Anzeige kann mit dem Button $> |U_{ss}| <$ ausgeblendet werden.

Folgende Aufnahmeparameter sind fest eingestellt und können nur im Quellprogramm geändert werden: Abtastfrequenz: $f_a = 44,1$ kHz ; Kanäle: Mono ; Auflösung: 16 bit.

- Ablauf der Aufnahme einer Sprachprobe:

a) Eingabe der Matrikelnummer (Student) oder der Telefonnummer (Mitarbeiter).

b) Auswahl des zu artikulierenden Vokals (Kontrastkonsonant | d |). Der gewählte Vokal wird auf der linken Seite des Bildes 1 (blinkt grün unterlegt) angezeigt.

c) Aufnahme mit den Button > START < und > STOPP <.

d) Eingabe des gemessenen Schallpegels in dB. Alle weiteren Funktionen des Programms sind ohne diese Eingabe blockiert (rote LED). Nach der Eingabe wird das Programm freigeschaltet (blaue LED).

e) Speicherung der WAV-Datei mit dem Schalter > Speichern < (Ordner: C:\Forsch). Der Dateiname wird automatisch generiert (Uhrzeit_Matrikel_Vokal_Pegel) und unter Pfad angezeigt.

f) Bei einer misslungenen Sprachprobe wird die Datei nicht abgespeichert, der Schalter > Löschen < betätigt und der Vorgang ab Punkt c) wiederholt.

g) Das rechts neben der Zeitfunktion eingeblendete Array zeigt die abgespeicherten Sprachproben an und kann mit dem Button > Save < ausgeblendet werden.

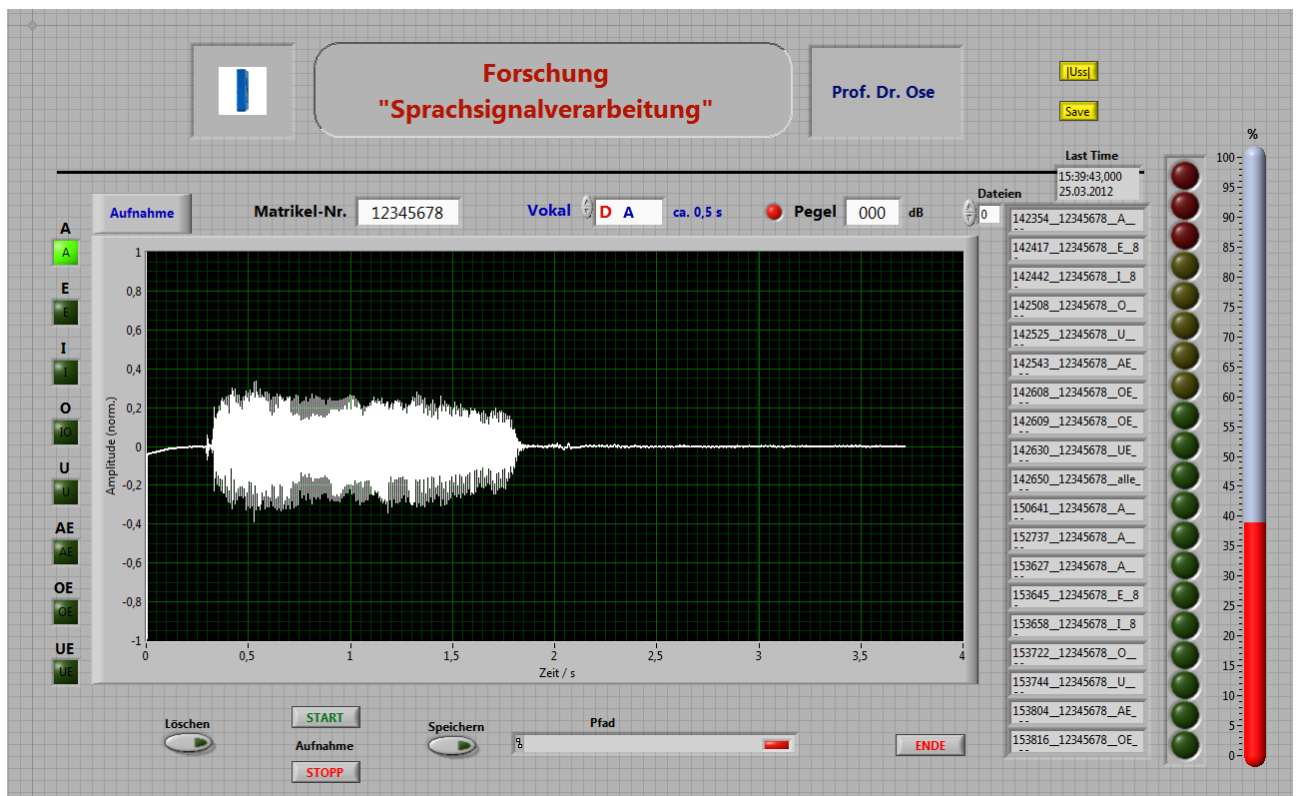


Bild 1: Bedienoberfläche zur Aufnahme einer Sprachprobe

Von jedem Sprecher werden 10 Dateien erzeugt:

* alle acht deutschen Langvokale einzeln mit dem Kontrastkonsonanten | d | (z.B.: | da |)

* alle acht deutschen Langvokale in der Folge mit Kontrast:
(| da | - | de | - | di | - | do | - | du | - | dä | - | dö | - | dü |)

* alle acht deutschen Langvokale in der Folge ohne Kontrast:
(| a | - | e | - | i | - | o | - | u | - | ä | - | ö | - | ü |)

Wenn alle WAV-Dateien eines Sprechers aufgenommen wurden, wird manuell ein Unterordner angelegt (Name = Matrikel-Nr.), in den dann die WAV-Dateien manuell verschoben werden.

3 Entwicklung eines Experimentiersystems

[Gliederung](#)

Das humane Perzeptionssystem beurteilt und bewertet Lautsprache mit subjektiven Empfindungsgrößen. Dabei wird aus dem Sprachsignal ein n -dimensionaler Merkmalsatz extrahiert, der die Grundlage für die nachfolgende Klassifikation bildet. Im Falle der Vokalperzeption leitet das humane Informationsverarbeitungssystem aus der Lautheits-Tonheits-Verteilung einen 21-dimensionalen Merkmalsatz ab. Dabei handelt es sich um die spezifischen Lautheiten in den ersten 21 Frequenzgruppen, die das Gehör über dem hörbaren Frequenzbereich aufspannt [5].

Das zu konzipierende Experimentiersystem hat demzufolge die Aufgabe, aus dem abgespeicherten Sprachsignal eine vokalische Phase zu selektieren und für diesen Signalabschnitt die Lautheits-Tonheits-Verteilung zu berechnen. Dieser Merkmalsatz bildet dann die Grundlage für die angestrebte Qualitätsbeurteilung.

3.1 Hauptprogramm

[Gliederung](#)

Bild 2 zeigt die Arbeitsoberfläche des Hauptprogramms. Hier wird zunächst die zu analysierende Sprachprobe geladen und als Zeitfunktion angezeigt. Mit dem Schalter > PLAY < ist eine Wiedergabe der Sprachprobe möglich.

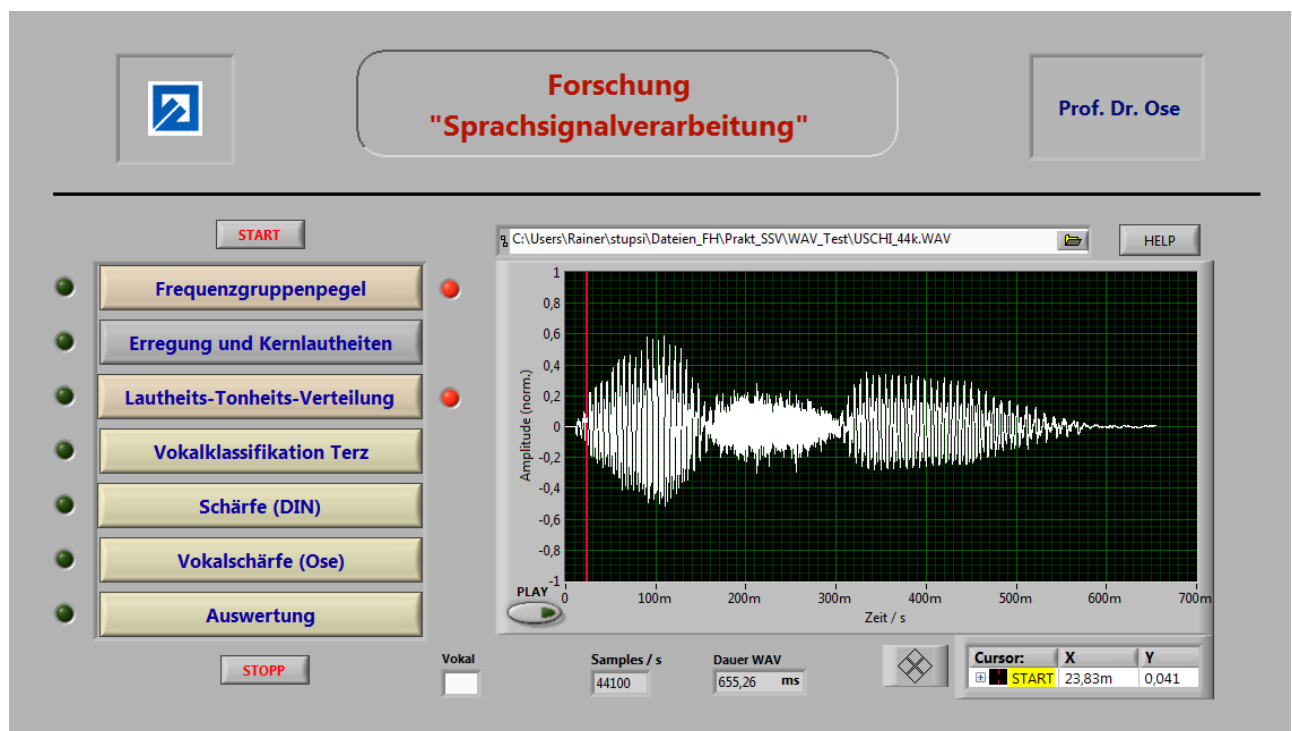


Bild 2: Arbeitsoberfläche zum Laden einer Sprachprobe (zugleich: Hauptprogramm)

Der START-Cursor (rot) ist frei verschiebbar und dient zur groben Einstellung des Startzeitpunktes des Analyse-Zeitfensters. Bei einer Vokalfolge ist der aktuelle Name des Vokals in das Eingabefenster „Vokal“ einzutragen. Von dieser Arbeitsoberfläche gelangt man in die einzelnen Unterprogramme, die zunächst in einer definierten Folge abzarbeiten sind. Bei der Umstellung auf ein Analysesystem, mit dem viele Datensätze zu bearbeiten sind, muss dieser Vorgang automatisiert werden. Bei einem Experimentiersystem ist es aber wünschenswert, jeden Bearbeitungsschritt einzeln auszuführen und die Zwischenergebnisse individuell zu bewerten. Nach Abarbeitung eines Unterprogramms erfolgt immer ein Rücksprung zur Arbeitsoberfläche des Hauptprogramms (Bild 2).

3.2 Bestimmung von Frequenzgruppenpegeln

[Gliederung](#)

Durch die Betätigung des Buttons > Frequenzgruppenpegel < (Bild 2 – oben) öffnet sich die im Bild 3 dargestellte Arbeitsoberfläche. Das dafür verantwortliche Unterprogramm führt eine Transformation des zu analysierenden Signals vom Zeitbereich in den Frequenzbereich durch.

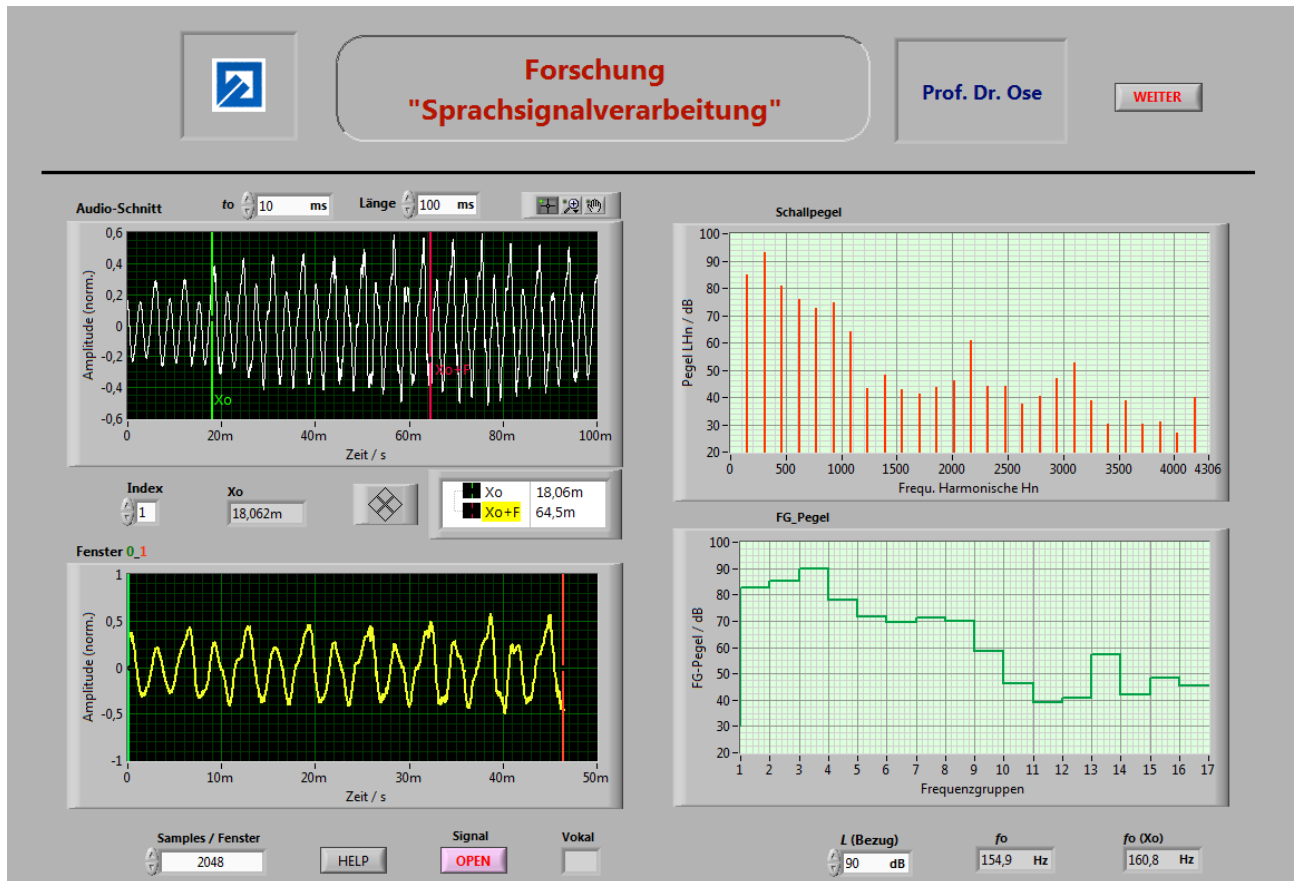


Bild 3: Arbeitsoberfläche zur Bestimmung von Frequenzgruppenpegeln

3.2.1 Festlegung des Analyse-Zeitfensters

[Gliederung](#)

Zur Durchführung dieser Transformation muss zunächst ein optimales Analyse-Zeitfenster auf das Signal gelegt werden. Durch Festlegung des Startzeitpunktes t_0 wird eine eventuell existierende Pause am Signalanfang sowie der Kontrastkonsonant entfernt. Die Eingabe „Länge“ bestimmt die Dauer des zu analysierenden Signalabschnitts (siehe Bild 3 – oben links – Audio-Schnitt). Mit den Cursors X_0 und X_0+F kann nun das Analyse-Zeitfenster festgelegt werden. Die Breite dieses Fensters (Abstand zwischen den beiden Cursors) wird mit dem Eingabewert „Samples / Fenster“ wie folgt berechnet:

$$\text{Fensterbreite} = \frac{\text{Samples / Fenster}}{\text{Abtastfrequenz}} \quad \text{bzw.:} \quad t_F = \frac{2^n}{f_a} \quad (1)$$

Bild 3 (unten links) zeigt die Zeitfunktion des Analyse-Zeitfensters.

Die nachfolgende Transformation beruht nun auf der Annahme, dass sich die Signalform aller vom Analyse-Zeitfenster erfassten Perioden nicht verändert. Bild 3 (links) zeigt, dass diese Annahme nur in quasistationären Abschnitten einer vokalischen Phase haltbar ist.

Zur automatischen Selektion solcher Abschnitte wurde eine Autokorrelationsfunktion (AKF) einprogrammiert. Sie ermöglicht die Bestimmung von Zeitpunkten, zu denen das Signal zu sich selbst die größte Ähnlichkeit aufweist. Diese Zeitpunkte werden abgespeichert und können über „Index“ nacheinander aufgerufen werden. Dieser Aufruf bewirkt eine neue Positionierung von X_0 sowie X_0+F und somit eine Verschiebung des Analyse-Zeitfensters im geschnittenen Signal „Audio-Schnitt“.

Bild 4 zeigt die prinzipielle Arbeitsweise dieser AKF. Dieses Bild kann mit dem Button > HELP < (siehe Bild 3) ein- und ausgeblendet werden.

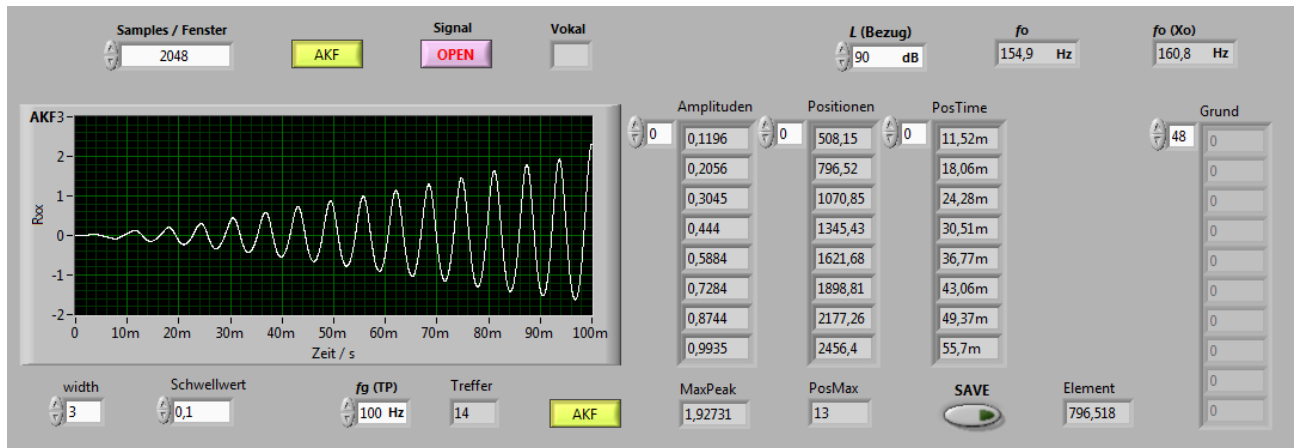


Bild 4: Visualisierung der Arbeitsweise der Autokorrelationsfunktion

Die Peaks in der AKF signalisieren die Zeitpunkte der größten Signalähnlichkeit und bestimmen den Anfangszeitpunkt X_0 eines Analyse-Zeitfensters. Ein einstellbarer Schwellwert sorgt dafür, dass zufällige Ähnlichkeiten unterdrückt werden. Die Anzeige „Treffer“ sagt aus, wie viele unterschiedliche Startzeitpunkte für das Analyse-Zeitfenster möglich sind. Bei einer zu groß gewählten Fensterbreite kann es allerdings passieren, dass das geschnittene Signal sein Ende erreicht und für das Fenster nicht mehr genügend Platz zur Verfügung steht.

Die Ergebnisse einer AKF können bei stimmhaften Sprachsignalen auch zur Bestimmung des Verlaufes der Grundfrequenz verwendet werden.

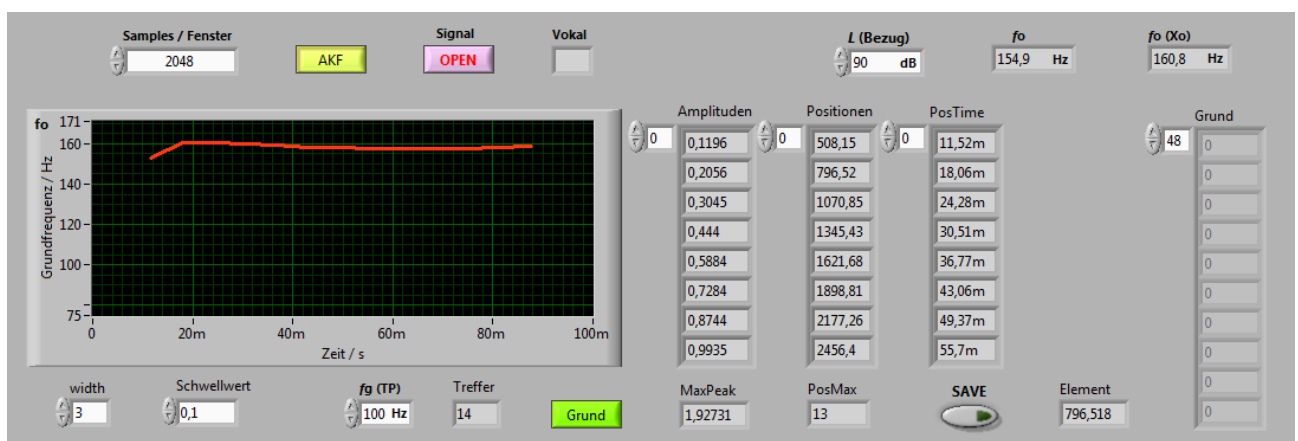


Bild 5: Visualisierung des Grundfrequenzverlaufs

Zur Darstellung dieses Verlaufes muss im Bild 4 der Button > AKF < betätigt werden. Damit wird zwischen der Darstellung des Verlaufes der AKF und der Grundfrequenz (Bild 5) umgeschaltet. Die zeitlichen Distanzen zwischen zwei benachbarten Elementen im Array „PosTime“ entsprechen der

jeweiligen Dauer einer betrachteten Periode und können über den Kehrwert in die Grundfrequenz des aktuellen Signalzeitpunktes umgerechnet werden.

Nach der Festlegung der Position und der Breite des Analyse-Zeitfensters werden nun für das gefensterete Signal die spektralen Komponenten bestimmt.

Im Sinne der Entwicklung eines Experimentiersystems kommen zwei parallel laufende Verfahren zum Einsatz:

- 1) Schnelle FOURIER-Transformation (FFT)
- 2) Parallel-Filterbank (20 Kanäle)

3.2.2 Schnelle FOURIER-Transformation

Gliederung

Bild 3 (oben rechts) zeigt das Ergebnis einer FFT. Das dazu verwendete SubVI berechnet das gemittelte einseitige Leistungsspektrum des Zeitsignals [7]. Daraus wird dann nach den bekannten Regeln das Amplitudenspektrum ermittelt. Dabei handelt es sich vorerst noch um Dämpfungen, die als eine Folge von Stützwerten $A_{St}(i)$ im Abstand Δf dargestellt werden.

$$\text{Abstand Stützwerte} = \frac{\text{Abtastfrequenz}}{\text{Samples / Fenster}} \quad \text{bzw.:} \quad \Delta f = \frac{f_a}{2^n} \quad (2)$$

Der spektralen Komponente mit der geringsten Dämpfung A_{\min} wird nun der gemessene Schallpegel L_{Bezug} (vgl. Kap. 2) zugeordnet. Damit wird jeder Stützwert als Schallpegel $L_{St}(i)$ dargestellt. Es gilt:

$$L_{St}(i) = A_{St}(i) + L_{\text{Bezug}} + |A_{\min}| \quad (3)$$

Der Bezugspegel (abgespeichert im Dateinamen) muss vom Bearbeiter im Feld „L(Bezug)“ eingegeben werden (vgl. Bild 3). Aus dem Spektrum der Stützwerte kann jetzt das Spektrum der Harmonischen als Linienspektrum $L_{Hn} = f(n_H)$ berechnet werden. Dazu wird die Grundfrequenz benötigt. Sie wurde im Ergebnis der AKF bereits ermittelt (Bild 5) und kann als Ausgabewert „ $f_0(Xo)$ “ (siehe Bild 3 – unten rechts) für die erste Periode rechts vom Cursor Xo abgelesen werden. Außerdem wird die mittlere Grundfrequenz über dem Analyse-Zeitfenster mit einer Tiefpass-Filterung bestimmt und als Ausgabewert „ f_0 “ (siehe Bild 3 – unten rechts) angezeigt. Ein Vergleich dieser beiden Ausgabewerte liefert dem Bearbeiter zugleich Informationen zur optimalen Lage des Analyse-Zeitfensters. Es gilt die Annahme, dass ein quasistationärer Abschnitt einer vokalischen Phase gefunden wurde, wenn beide Ausgabewerte nahezu gleich sind.

Die Frequenzlage der n -ten Harmonischen (n_H) ergibt sich dann wie folgt:

$$f_{Hn} = n_H \cdot f_0 \quad (4)$$

Für diese Frequenz wird nun der nächstgelegene Stützwert gesucht und mit seinen beiden Nachbarn zum Pegel der Harmonischen überlagert:

$$L_{Hn} = 10 \cdot \lg \sum_{(i)} 10^{0,1 \cdot L_i} \quad \text{mit:} \quad (n_{St} - 1) \leq i \leq (n_{St} + 1) \quad (5)$$

Bild 3 zeigt oben rechts das erzielte Ergebnis in Form eines Linienspektrums. Es werden die Schallpegel der Harmonischen als Funktion der Frequenzlage der Harmonischen dargestellt. Zur Berechnung der Harmonischen aus den Stützwerten wird die mit der Tiefpass-Filterung ermittelte Grundfrequenz in Gleich. (4) verwendet.

3.2.3 Parallelfilterung

Gliederung

Das zweite Verfahren arbeitet auf der Grundlage einer Parallelfilterbank. Dazu werden 20 Filter (Butterworth – 3. Ordnung) parallel geschaltet. Die Grenzfrequenzen der Filter entsprechen den Grenzfrequenzen der Frequenzgruppen des menschlichen Gehörs (siehe Anlage 1 – Tabelle 1). Auf die Nachbildung der Frequenzgruppen Nr. 21 bis Nr. 24 wurde verzichtet, da sie im Rahmen der Vokalperzeption keine relevanten Informationen mehr liefern.

Dem Filter mit der geringsten Dämpfung A_{\min} wird der gemessene Schallpegel L_{Bezug} (vgl. Kap. 2) zugeordnet. Damit wird die Dämpfung jedes Filters als Schallpegel $L_G(j)$ dargestellt. Es gilt:

$$L_G(j) = A_G(j) + L_{\text{Bezug}} + |A_{\min}| \quad (6)$$

Der Bezugspegel (abgespeichert im Dateinamen) muss vom Bearbeiter im Feld „ $L(\text{Bezug})$ “ eingegeben werden (vgl. Bild 3).

Im Ergebnis erhält man nicht mehr das Spektrum der einzelnen Harmonischen, sondern den resultierenden Pegel in den einzelnen Frequenzgruppen $L_G(j)$. Bild 3 (unten rechts) zeigt ein Beispiel. Es werden die Schallpegel in den ersten 17 Frequenzgruppen als Funktion der oberen Grenztöne z_{go} (zugleich Nr. der jeweiligen Frequenzgruppe) dargestellt.

Die unter Abschnitt 3.2 beschriebenen Programmkomponenten wurden zu einem Unterprogramm `Pegel_EXP_AKF.vi` zusammengefasst. Dieses Unterprogramm wird so lange in einer While-Schleife abgearbeitet, bis man den Button `> Weiter <` betätigt. Die bei jedem Schleifendurchlauf berechneten Datensätze werden in Arrays abgelegt und stehen für die Weiterverarbeitung als globale Variable zur Verfügung:

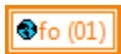


Datensatz zur Darstellung des Linienspektrums $L_{Hn} = f(f_{Hn})$



Datensatz zur Darstellung der Frequenzgruppenpegel $L_G = f(z)$

Außerdem werden folgende globale Variable gesetzt:



Mittlere Grundfrequenz im Analyse-Zeitfenster



Aktuell untersuchter Vokal

Der Button `> Weiter <` bewirkt einen Rücksprung in das Hauptprogramm `Haupt_EXPERT.vi`. Dort wird die Kontrolllampe links vom Button `> Frequenzgruppenpegel <` auf ‚Grün‘ geschaltet, um dem Bearbeiter zu signalisieren, dass diese Programmschritt abgearbeitet ist. Die Kontrolllampe rechts vom Button `> Frequenzgruppenpegel <` schaltet von ‚Rot‘ auf ‚Blau‘ und zeigt damit an, dass die nachfolgenden zwei Bearbeitungsschritte freigegeben sind. Die restlichen Bearbeitungsschritte sind noch gesperrt, bis die Daten für die Lautheits-Tonheits-Verteilung vorliegen. Diese Daten können direkt über den Button `> Lautheits-Tonheits-Verteilung <` mit dem unter Abschnitt 3.4 beschriebenen Unterprogramm oder über den Zwischenschritt `> Erregung und Kernlautheiten <` (vgl. Abschnitt 3.3) ermittelt werden. Wenn man auf den Zwischenschritt `> Erregung und Kernlautheiten <` verzichtet, werden diese Daten automatisch ermittelt und für die Berechnung der Lautheits-Tonheits-Verteilung ohne Anzeige über einen globalen Parameter bereitgestellt.

3.3 Berechnung von Erregungspegeln und Kernlautheiten

[Gliederung](#)

Zur Bestimmung der Lautheits-Tonheits-Verteilung müssen zunächst die Erregungspegel aus den Frequenzgruppenpegeln berechnet werden. Die Frequenzgruppenpegel L_G repräsentieren noch physikalische Größen (messbar z.B. als Ausgangsgrößen von Filtern – vgl. Abschn. 3.2).

Die Erregungspegel L_E erfassen zusätzlich das Dämpfungsmaß a_0 des menschlichen Gehörs sowie Unterschiede zwischen einem ebenen und einem diffusen Schallfeld. Für den Erregungspegel der Frequenzgruppe j gilt für ein ebenes Schallfeld:

$$L_E(j) = L_G(j) - a_0(j) \quad (7)$$

Bei einem diffusen Schallfeld muss zusätzlich die Pegeldifferenz zwischen ebenem und diffussem Schallfeld ΔL_{ED} berücksichtigt werden:

$$L_E(j) = L_G(j) - a_0(j) + \Delta L_{ED}(j) \quad (8)$$

Die Größen $a_0(j)$ und $\Delta L_{ED}(j)$ sind in der Anlage 1 aufgelistet. Dabei ist zu anmerken, dass das Dämpfungsmaß a_0 des Gehörs in [5] und [6] mit einem Kleinbuchstaben gekennzeichnet wird. Dieses Dämpfungsmaß ist aber nicht von der Zeit abhängig (Alterungserscheinungen werden hier nicht berücksichtigt).

Bild 6 zeigt auf der linken Seite den Verlauf des Erregungspegel (hier: diffuses Schallfeld) als Funktion der Tonheit. Die Tonheit z (Einheit: Bark) ist die zur Frequenz psychisch relevante Größe. Jede Frequenzgruppe hat eine konstante Breite von 1 Bark, obwohl die physikalische Bandbreite B mit ansteigender Tonheit zunimmt.

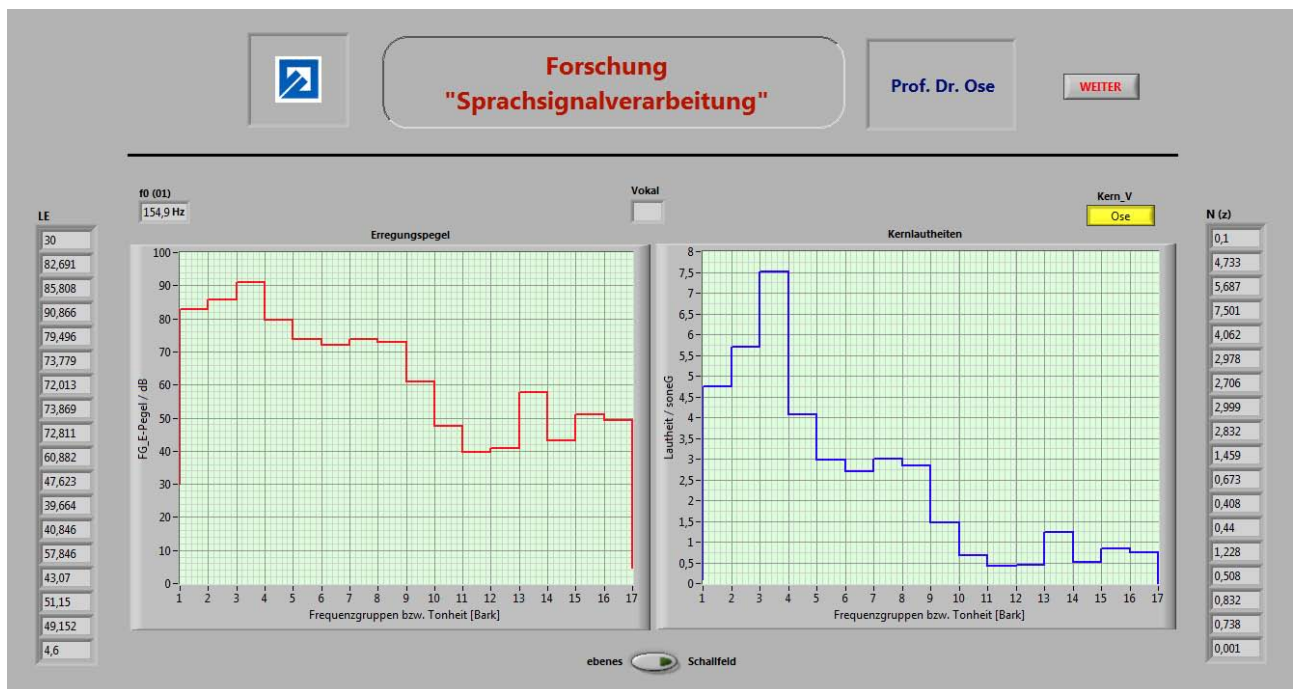


Bild 6: Erregungspegel und Kernlautheiten in den einzelnen Frequenzgruppen

Mit dem Schalter „ebenes Schallfeld“ kann die Berechnungsvorschrift für den Erregungspegel von Gleich. (6) auf Gleich. (7) umgeschaltet werden.

Aus den Erregungspegeln werden nun die Kernlautheiten in den einzelnen Frequenzgruppen berechnet. Nach [6] gilt (ohne indizierte Schreibweise):

$$N'_{\text{Kern}} = 0,064 \cdot 10^{0,025 \cdot L_{\text{EHS}}} \cdot \{ [1 + 0,25 \cdot 10^{0,1 \cdot (L_{\text{E}} - L_{\text{EHS}})}]^{0,25} - 1 \} \frac{\text{sone}_G}{\text{Bark}} \quad (9)$$

L_{EHS} beschreibt dabei den Erregungspegel an der Ruhehörschwelle (vgl. Tabellen 2 und 3).

Nach [4] gilt (ohne indizierte Schreibweise):

$$N'_{\text{Kern}} = 0,08 \cdot 10^{0,023 \cdot L_{\text{EHS}}} \cdot \{ [0,5 + 0,5 \cdot 10^{0,1 \cdot (L_{\text{E}} - L_{\text{EHS}})}]^{0,23} - 1 \} \frac{\text{sone}_G}{\text{Bark}} \quad (10)$$

Nach DIN 45631 [10] gilt (ohne indizierte Schreibweise):

$$N'_{\text{Kern}} = 0,0635 \cdot 10^{0,025 \cdot L_{\text{EHS}}} \cdot \{ [1 - s + s \cdot 10^{0,1 \cdot (L_{\text{E}} - L_{\text{EHS}})}]^{0,25} - 1 \} \frac{\text{sone}_G}{\text{Bark}} \quad (11)$$

Im Ergebnis umfangreicher Experimente konnte festgestellt werden, dass sich die Ergebnisse dieser Berechnungsvorschriften nur unwesentlich voneinander unterscheiden. Der Schwellenfaktor s in der Gleich. (11) wird in [10] mit $s = (0,25 \dots 0,3)$ angenommen. Er dient zum Abgleich der berechneten Gesamtlautheit auf den definierten Wert $N_{\text{ges}} = 1$ sone, wenn ein sinusförmiges Testsignal ($f = 1$ kHz) mit einem Pegel von 40 dB verwendet wird.

Wenn man die unterschiedlichen Zahlenwerte durch allgemeine Konstanten ausdrückt, kann eine variable Berechnungsvorschrift im Sinne eines Experimentiersystems programmiert werden:

$$N'_{\text{Kern}}(j) = K_1 \cdot 10^{K_2 \cdot L_{\text{EHS}}(j)} \cdot \{ [K_3 + K_4 \cdot 10^{0,1 \cdot (L_{\text{E}}(j) - L_{\text{EHS}}(j))}]^{K_5} - 1 \} \quad (12)$$

Die verschiedenen Zahlenwerte wurden in einer umschaltbaren Case-Struktur abgelegt und werden (je nach gewählter Berechnungsvorschrift) über den Button > Kern_V < (siehe Bild 6 – oben rechts) abgerufen. Die Kernlautheit wird nach Gleich. (12) mit den Konstanten aus [4] oder aus [10] berechnet. Im Abschnitt 3.4 werden die jeweiligen Zahlenwerte optimiert.

Die im Abschnitt 3.3 beschriebenen Programmkomponenten wurden zu einem Unterprogramm Kern_EXP.vi zusammengefasst. Dieses Unterprogramm wird so lange in einer While-Schleife abgearbeitet, bis man den Button > Weiter < betätigt. Die bei jedem Schleifendurchlauf berechneten Datensätze werden in Arrays abgelegt und stehen für die Weiterverarbeitung als globale Variable zur Verfügung:



Datensatz zur Darstellung der Erregungspegel $L_{\text{E}} = f(z)$



Datensatz zur Darstellung der Kernlautheiten $N'_{\text{Kern}} = f(z)$

3.4 Berechnung der Lautheits-Tonheits-Verteilung

[Gliederung](#)

Jede Kernlautheit bewirkt in den nachfolgenden Frequenzgruppen eine Flankenlautheit, die mit zunehmender Tonheit nach einer Hyperbelfunktion abklingt. Eine solche Flanke kann Kernlautheiten in nachfolgenden Frequenzgruppen vollständig oder teilweise überdecken (Effekt der Verdeckung). Die jeweils wirksame Lautheit (Kern / Flanke / Kombination von beiden) wird als spezifische Lautheit pro Frequenzgruppe $N'(z)$ bezeichnet und mit der Einheit sone_G / Bark angegeben. Bild 7 zeigt ein Beispiel, das zum Test der entwickelten LabVIEW-Software verwendet wurde.

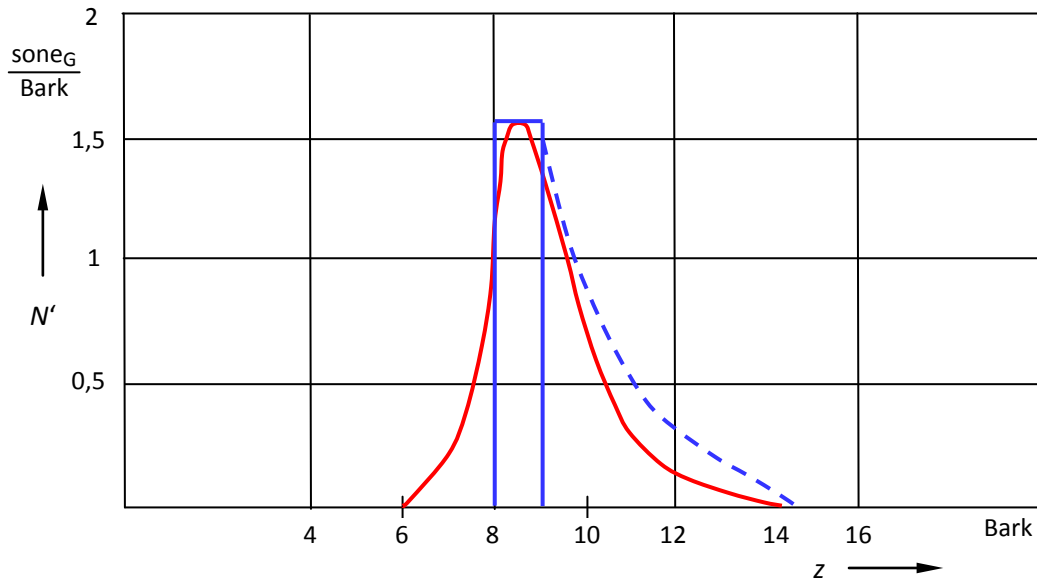


Bild 7: Lautheit eines 1-kHz-Tones mit einem Schallpegel von 60 dB als Funktion der Tonheit (rot)

(blau: zweckmäßige Näherung – Kern durchgezogen / Flanke gestrichelt)

Quelle: [5] und [6]

Die Gesamtlautheit einer akustischen Erscheinung ergibt sich dann wie folgt:

$$N_{\text{ges}} = \int_{z=0}^{24 \text{ Bark}} N'(z) \cdot dz \quad (13)$$

Bild 8 zeigt die Arbeitsoberfläche eines Testprogramms (Haupteinsatzgebiet: Praktikum Sprachsignalverarbeitung), mit dem die Berechnung der Gesamtlautheit mit eingegebenen Frequenzgruppenpegel möglich ist (Test 1 = Lautheit gemäß Definition [5] und Test 2 = Beispielwert aus [10]).

Test 1: $L_G = 40 \text{ dB}$ (K 9: $f_m = 1 \text{ kHz}$)

nach [4]: $N_{\text{ges}} = 1,0013 \text{ sone}$

$L_G = 40 \text{ dB}$ (K 9: $f_m = 1 \text{ kHz}$)

nach [10]: $N_{\text{ges}} = 1,0058 \text{ sone}$

Soll: 1 sone

Test 2: $L_G = 70 \text{ dB}$ (K 9: $f_m = 1 \text{ kHz}$)

nach [4]: $N_{\text{ges}} = 7,12 \text{ sone}$

$L_G = 70 \text{ dB}$ (K 9: $f_m = 1 \text{ kHz}$)

nach [10]: $N_{\text{ges}} = 7,99 \text{ sone}$

Soll: 8 sone

Für die Konstanten in Gleich. (12) gilt dann:

Nach [4]: $K_1 = 0,08$ $K_2 = 0,023$ $K_3 = 0,77$ $K_4 = s = 0,23$ $K_5 = 0,23$

Nach [10]: $K_1 = 0,0635$ $K_2 = 0,025$ $K_3 = 0,73$ $K_4 = s = 0,27$ $K_5 = 0,25$

Die Berechnung der Gesamtlautheit nach [10] scheint genauer zu sein. Es ist aber zu beachten, dass in [10] mit Terzpegeln und in [4] mit Frequenzgruppenpegeln gearbeitet wurde.

Es werden demzufolge beide Berechnungsvorschriften (umschaltbar) für die weiteren Analysen zur Verfügung stehen.

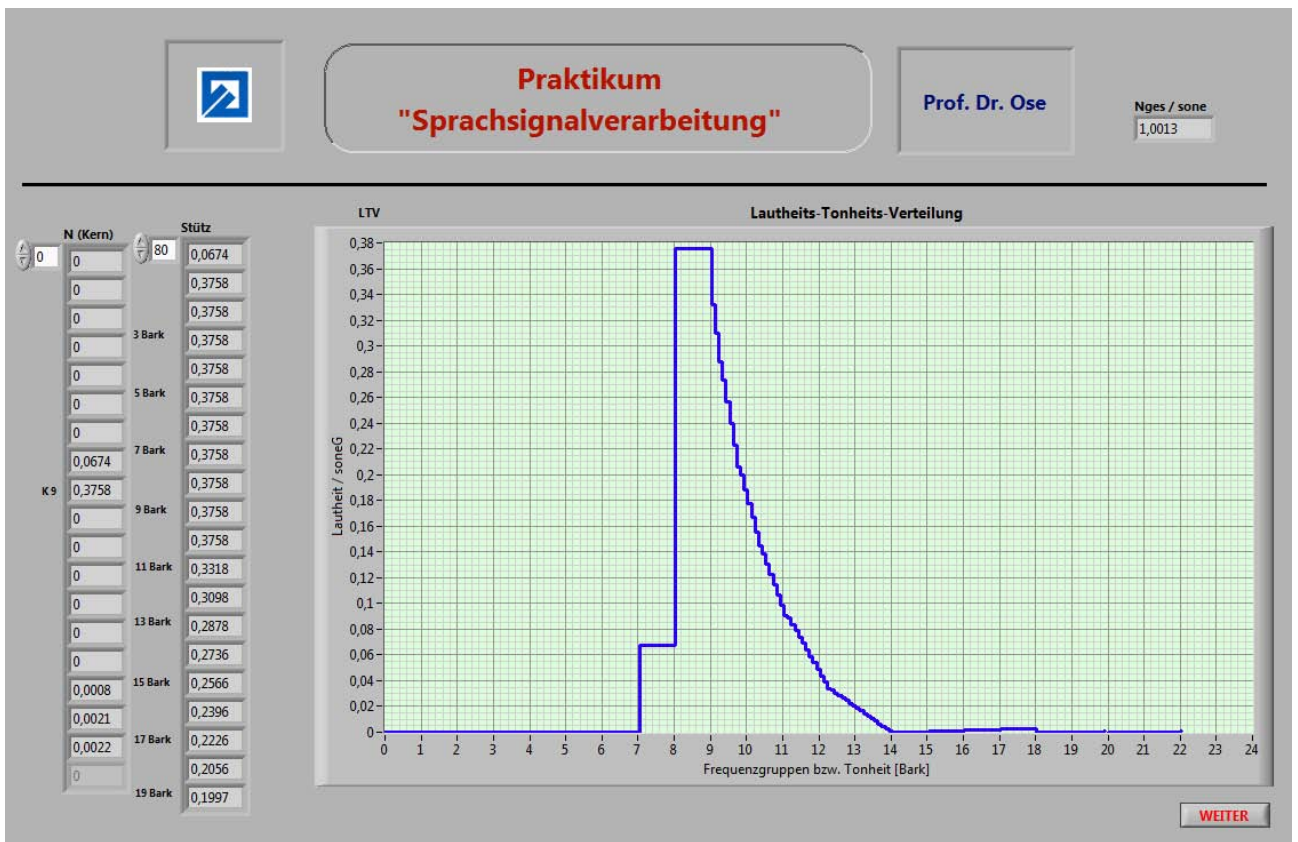
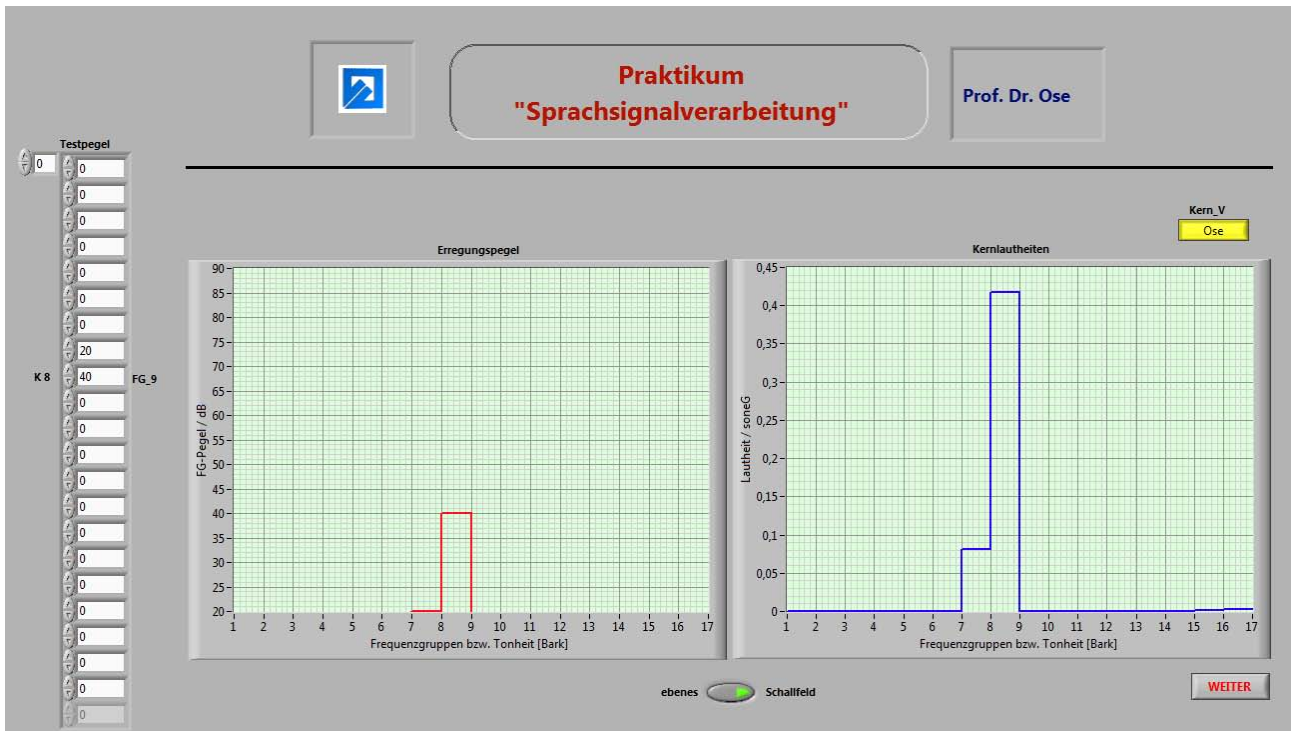


Bild 8: Lautheit eines 1-kHz-Tones mit einem Schallpegel von 40 dB als Funktion der Tonheit
 oben: Erregungspegel und Kernlautheit
 unten: Lautheits-Tonheits-Verteilung

Für die Berechnung der Flankenlautheiten wird der in Anlage 2 dargestellte Algorithmus verwendet. Nach der Berechnung der Kernlautheiten (siehe Bild 8 – oben) werden in allen nachfolgenden Frequenzgruppen jeweils 10 Stützwerte bestimmt. Im Bild 8 (unten) kann man diese Stützwerte deutlich erkennen. Aus diesem Grund wurde im Bild 8 auf eine Glättung des Kurvenverlaufs verzichtet.

Für die weiteren Verarbeitungsschritte werden die spezifischen Lautheiten pro Frequenzgruppe benötigt. Bei einer dominanten Kernlautheit ist dieser Wert für die jeweilige Frequenzgruppe aus Abschnitt 3.3 (vgl. auch Bild 8) bereits bekannt. Bei einer teilweisen oder vollständigen Verdeckung werden die 10 Stützwerte pro Frequenzgruppe in einen mittleren Lautheitswert umgerechnet. Man erhält die im Bild 9 dargestellte Verteilung (Beispiel: Vokal | i |).

Das humane Perzeptionssystem teilt die Tonheitsskala zunächst in zwei Bereiche auf. Die Grenze liegt bei 10,5 Bark (ca. 1,4 kHz). In beiden Tonheitsbereichen bilden sich in der Lautheits-Tonheits-Verteilung lokale Maxima aus. Sie können (müssen aber nicht) erkenntnisspezifisch sein.

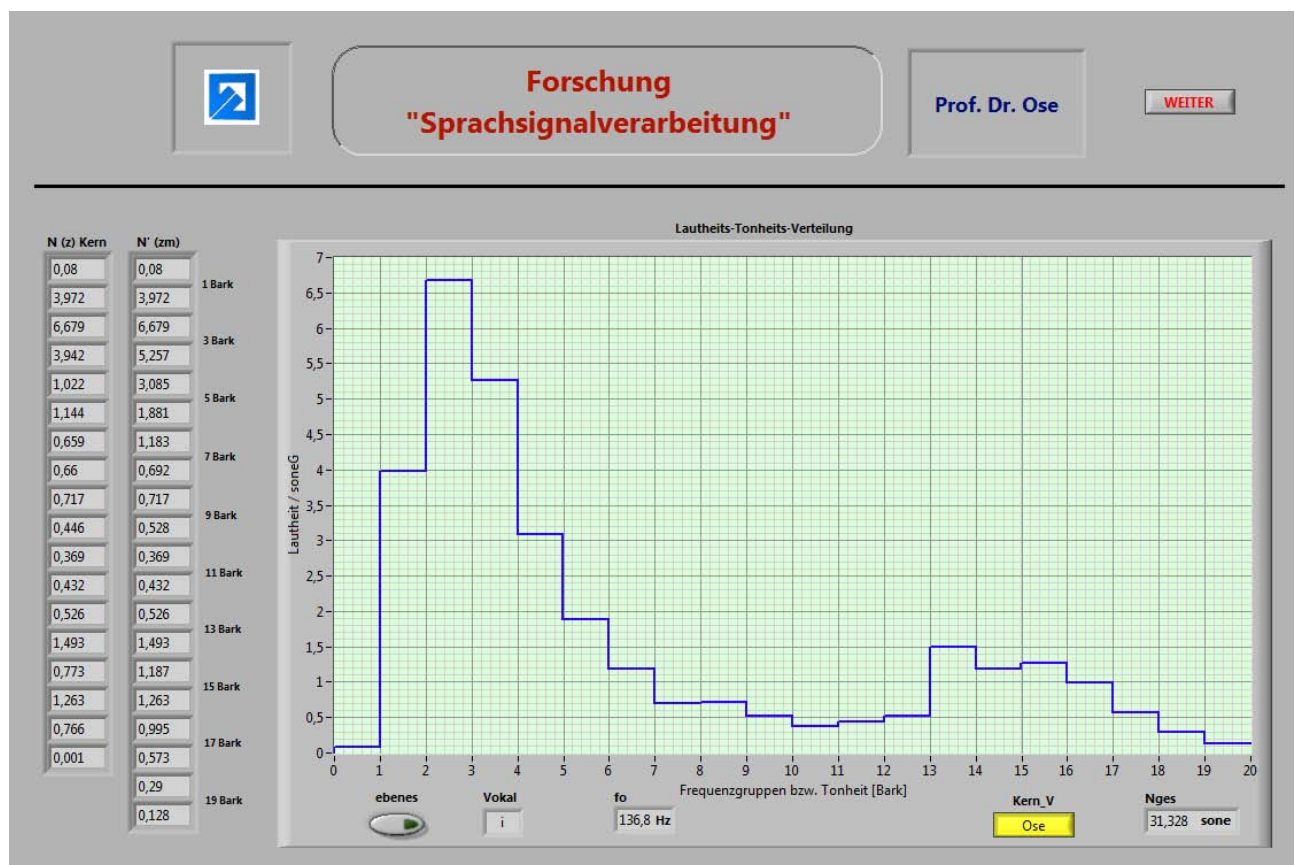


Bild 9: Lautheits-Tonheits-Verteilung des Vokals | i | aus „Uschi“

Die im Abschnitt 3.4 beschriebenen Programmkomponenten wurden zu einem Unterprogramm Flanken_EXP.vi zusammengefasst. Dieses Unterprogramm wird so lange in einer While-Schleife abgearbeitet, bis man den Button > Weiter < betätigt. Der bei jedem Schleifendurchlauf berechnete Datensatz wird in einem Array abgelegt und steht für die Weiterverarbeitung als globale Variable zur Verfügung:



Datensatz zur Darstellung der spezifischen Lautheiten $N' = f(z)$

Auf der Grundlage dieses Datensatzes können nun Vokalpräsenzen berechnet werden, die eine Aussage zur subjektiven Zuordnung zu einer Eigenschaftsklasse / Vokalklasse ermöglichen.

3.5 Vokalklassifikation

[Gliederung](#)

Die Theorie zur Vokalperzeption wird in [4], [8] und [9] ausführlich beschrieben. An dieser Stelle sollen lediglich grundlegende Verarbeitungsprinzipien vorgestellt werden, die zum Verständnis der Arbeitsweise der Klassifikationssoftware (siehe Bild 10) erforderlich sind.

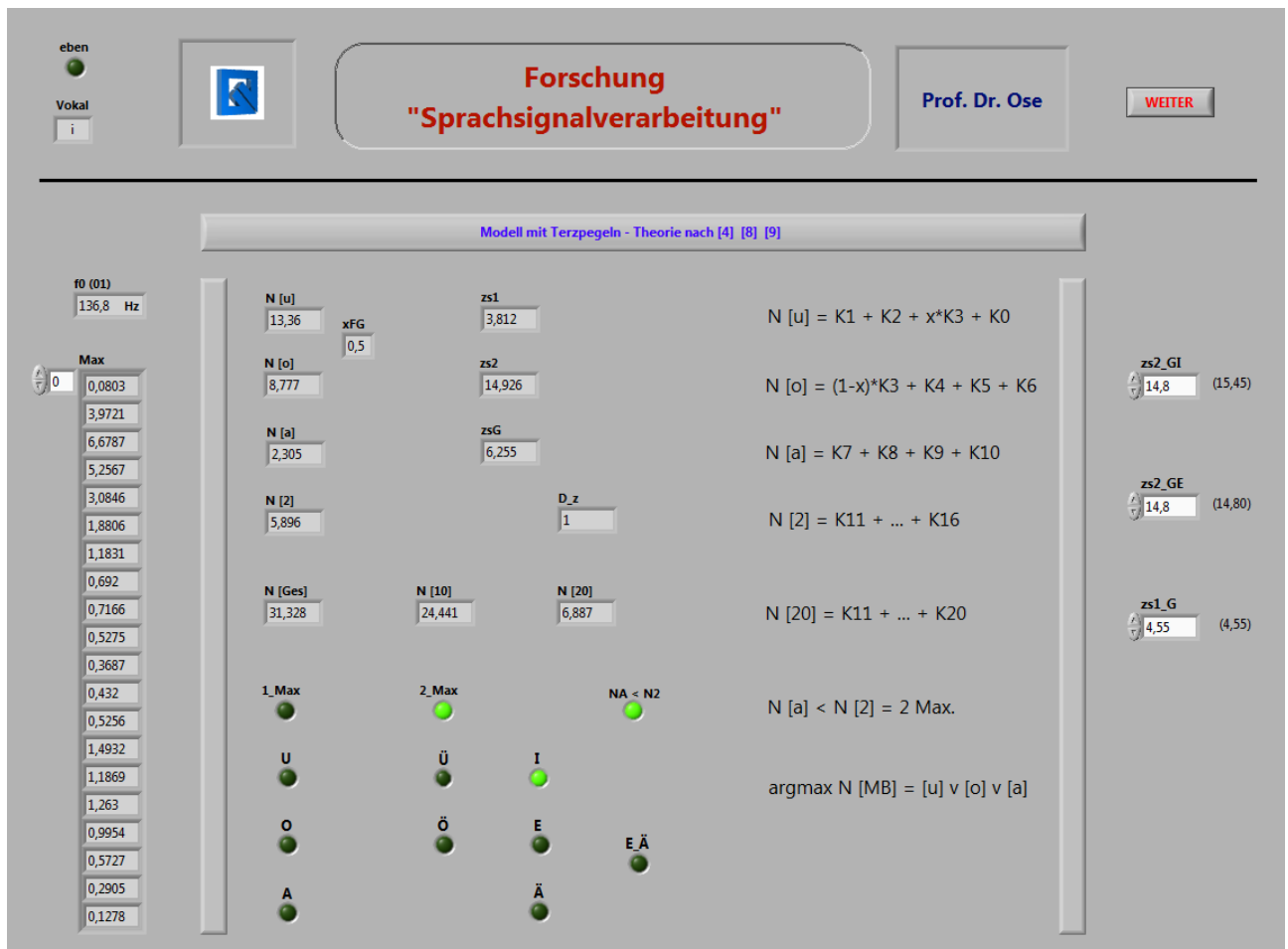


Bild 10: Arbeitsoberfläche des Programms zur Vokalklassifikation

Das humane Perzeptionssystem verarbeitet quasistationäre vokalische Schallereignisse nach zwei grundlegenden Prinzipien: • Dominanzprinzip • Positionsprinzip

Der dazu erforderliche Merkmalsatz wird aus der Lautheits-Tonheits-Verteilung abgeleitet und besteht aus den spezifischen Lautheiten der ersten 21 Frequenzgruppen. Die präzise Abbildung dieses Merkmalsatzes (hier: Genauigkeit der Berechnung – vgl. auch Kap. 5) bestimmt die Glaubwürdigkeit des Erkennungsergebnisses.

3.5.1 Dominanzprinzip

Aus den 21 Merkmalen werden Summenlautheiten in vokalspezifischen Merkmalbändern gebildet. Der humane Diskriminator führt eine erste Kompression des Merkmalsatzes durch, indem er den zu analysierenden Tonheitsbereich bei $z_a = 10,5$ Bark in zwei Teilbereiche aufteilt.

Der Tonheitsbereich unterhalb 10,5 Bark ($0 < z < z_a$) trägt Informationen zum ersten erkenntnis-spezifischen Maximum einer vokalischen Realisierung. Dieser Bereich wird von drei vokalspezi-fischen Merkmalbändern mit der Breite von jeweils $\Delta z = 3,5$ Bark ausgefüllt. Die oberen Grenzen

dieser Merkmalbänder sind von der Grundfrequenz der Realisierung [Gleich. (14)] sowie von der Lautstatistik (Aufretenshäufigkeit einer Klasse) und vom Kontrast (kurzfristige Vergangenheit des Perzeptionsresultates) abhängig. Die diesen Abhängigkeiten zugrunde liegenden Gesetzmäßigkeiten werden über die Bildung des Referenzwissens erworben. Es gilt:

$$f_{\text{MBo}} = f_0 \cdot \coth \frac{f_0}{f_{\text{go}}} \quad (14)$$

In diesen drei vokalspezifischen Merkmalbändern werden Summenlautheiten gebildet, die ein Maß für die Ähnlichkeit mit den Eigenschaften folgender Vokalgruppen darstellen:

$$[\text{u}]\text{-ähnlich: } 0 \dots 3,5 \text{ Bark } (f_{\text{go}} \approx 350 \text{ Hz}) \quad \Rightarrow \quad d_1(\vec{N})$$

$$[\text{o}]\text{-ähnlich: } 3,5 \dots 7,0 \text{ Bark } (f_{\text{go}} \approx 770 \text{ Hz}) \quad \Rightarrow \quad d_2(\vec{N})$$

$$[\text{a}]\text{-ähnlich: } 7 \dots 10,5 \text{ Bark } (f_{\text{go}} \approx 1,4 \text{ kHz}) \quad \Rightarrow \quad d_3(\vec{N})$$

Für die Bestimmung dieser Unterscheidungsfunktionen gilt:

$$[\text{u}] \text{ – Merkmalband: } \quad d_1(\vec{N}) = \sum_{z=0}^{3,5\text{Bark}} N'(z) \quad (15)$$

$$[\text{o}] \text{ – Merkmalband: } \quad d_2(\vec{N}) = \sum_{z=3,5\text{Bark}}^{7\text{Bark}} N'(z) \quad (16)$$

$$[\text{a}] \text{ – Merkmalband: } \quad d_3(\vec{N}) = \sum_{z=7\text{Bark}}^{10,5\text{Bark}} N'(z) \quad (17)$$

Der Tonheitsbereich oberhalb 10,5 Bark ($z_a \leq z \leq 21$ Bark) trägt Informationen zum zweiten erkenntungsspezifischen Maximum einer vokalischen Realisierung. Die in diesem Merkmalband gebildete Summenlautheit ist ein Maß für die Existenz von Eigenschaften, die den Vokalgruppen mit zwei erkenntungsspezifischen Maxima zugeordnet werden können:

$$> [\text{a}]\text{-ähnlich: } 10,5 \dots 21 \text{ Bark } (f_{\text{go}} \approx 7700 \text{ Hz}) \quad \Rightarrow \quad d_4(\vec{N})$$

Für die Bestimmung dieser Unterscheidungsfunktion gilt:

$$> [\text{a}] \text{ – Merkmalband: } \quad d_4(\vec{N}) = \sum_{z=10,5\text{Bark}}^{21\text{Bark}} N'(z) \quad (18)$$

Nun erfolgt ein Vergleich zwischen den Summenlautheiten der drei Merkmalbänder im Bereich unterhalb von $z_a = 10,5$ Bark. Die in ihrer Intensität dominierende Unterscheidungsfunktion $d_1(\vec{N})$, $d_2(\vec{N})$ oder $d_3(\vec{N})$ gibt die größte Ähnlichkeit zu einem der Vokale mit einem erkenntungsspezifischen Maximum ([u]-, [o]- oder [a]-ähnlich) an. Durch einen weiteren Vergleich zwischen der Summenlautheit oberhalb des [a]-Merkmalbandes und der Summenlautheit im [a]-Merkmalband wird ermittelt, ob es

sich um eine Konfiguration mit einem erkenntungsspezifischen Maximum oder mit zwei erkenntungsspezifischen Maxima handelt. Aus diesen Vergleichen werden zwei Eigenschaftsklassen abgeleitet:

$$e_{k1} = \operatorname{argmax}_{i=1}^3 d_i(\vec{N}) \quad (19)$$

$$e_{k2} = \operatorname{argmax}_{i=3}^4 d_i(\vec{N}) \quad (20)$$

3.5.2 Positionsprinzip

Mit der Bildung von Summenlautheiten in vokalspezifischen Merkmalbändern leitet das humane Perzeptionssystem subjektiv empfundene Intensitätsinformationen (Empfindungsgrößen) aus einer vokalischen Realisierung ab. Die Positionierung dieser Intensitäten wird über Tonheitsschwerpunkte ermittelt. Ein Tonheitsschwerpunkt beschreibt die Lage von Lautheitskonzentrationen auf der Tonheitsachse und kann als Maß für subjektiv empfundene Positionssinformation (Empfindungsgröße) aufgefasst werden. Diese Positionsinformationen leitet das humane Perzeptionssystem für die beiden Tonheitsbereiche unterhalb und oberhalb von z_a ab. Dazu wird jede spezifische Lautheit mit der Tonheit ihrer Frequenzgruppe bewertet. Die Summe dieser bewerteten Lautheitskomponenten wird dann auf die Summenlautheit des betrachteten Tonheitsbereiches bezogen:

$$d_5(\vec{N}) = z_{S1} = \frac{\sum_{z=0}^{10,5\text{Bark}} z \cdot N'(z)}{\sum_{z=0}^{10,5\text{Bark}} N'(z)} \quad (21)$$

$$d_6(\vec{N}) = z_{S2} = \frac{\sum_{z=10,5\text{Bark}}^{21\text{Bark}} z \cdot N'(z)}{\sum_{z=10,5\text{Bark}}^{21\text{Bark}} N'(z)} \quad (22)$$

Nun erfolgt ein Vergleich zwischen dem jeweiligen Tonheitsschwerpunkt z_S und einem bereichsspezifischen Grenzscherpunkt z_G . Aus der Lage eines Tonheitsschwerpunktes relativ zu seinem Grenzwert, der im Referenzwissen abgelegt ist, wird die größte Ähnlichkeit zu einem der Vokale mit zwei erkenntungsspezifischen Maxima ([ü] – [i] bzw. [ö] – [e] oder [e] – [ä]) ermittelt.

$$e_{k3} = \begin{cases} 0 & (z_{S2} \leq z_{G2}) \\ 1 & (z_{S2} > z_{G2}) \end{cases} \quad (23)$$

$$e_{k4} = \begin{cases} 0 & (z_{S1} \leq z_{G1}) \\ 1 & (z_{S1} > z_{G1}) \end{cases} \quad (24)$$

Der Grenzscherpunkt im oberen Tonheitsbereich $z \geq z_a$ liegt bei $z_{G2} = 14$ Bark.

Der Grenzscherpunkt im unteren Tonheitsbereich $z < z_a$ liegt bei $z_{G1} = 5,9$ Bark.

Tabelle 1 zeigt in einer Zusammenfassung die Ableitung von Eigenschaftsklassen aus den vom Diskriminator gebildeten Unterscheidungsfunktionen. Mit diesen Eigenschaftsklassen ist die in der rechten Spalte angegebene Vokalklassenzuordnung möglich.

Tabelle 1: Ableitung von Eigenschaftsklassen zur Vokalklassenzuordnung

e_{k1}	e_{k2}	e_{k3}	e_{k4}	[k]
$d_1(\vec{N}) = \max.$	$d_3(\vec{N}) = \max.$			[u]
$d_2(\vec{N}) = \max.$	$d_3(\vec{N}) = \max.$			[o]
$d_3(\vec{N}) = \max.$	$d_3(\vec{N}) = \max.$			[a]
$d_1(\vec{N}) = \max.$	$d_4(\vec{N}) = \max.$	0		[ü]
$d_1(\vec{N}) = \max.$	$d_4(\vec{N}) = \max.$	1		[i]
$d_2(\vec{N}) = \max.$	$d_4(\vec{N}) = \max.$	0		[ö]
$d_2(\vec{N}) = \max.$	$d_4(\vec{N}) = \max.$	1	0	[e]
$d_2 \vee d_3 = \max.$	$d_4(\vec{N}) = \max.$	1	1	[ä]

Aus dieser Tabelle ist ersichtlich, dass für Vokale mit einem erkenntungsspezifischen Maximum ([u], [o] und [a]) bereits einfache Verarbeitungsmechanismen (hier: Dominanz der Summenlautheit in einem der drei Merkmalsbänder unterhalb z_a) zu einem Erkennungsergebnis führen.

Vokale mit zwei erkenntungsspezifischen Maxima ([ü], [i], [ö], [e] und [ä]) repräsentieren eine höhere Reizqualität. Für ihre Klassenzuordnung müssen demzufolge weitere Verarbeitungsmechanismen aktiviert werden. Das gilt insbesondere für die Vokale [e] und [ä], die eine große Ähnlichkeit in ihrer spektralen Konfiguration aufweisen können. Die Aktivierung von Verarbeitungsmechanismen nimmt das humane Perzeptionssystem längs einer reizabhängigen Perzeptionshierarchie vor.

Die im Abschnitt 3.5 beschriebenen Programmkomponenten wurden zu einem Unterprogramm `Unt_Funkt_EXP_Terz.vi` zusammengefasst. Dieses Unterprogramm wird so lange in einer While-Schleife abgearbeitet, bis man den Button > Weiter < betätigt. Der bei jedem Schleifendurchlauf berechnete Datensatz wird in einem Array abgelegt und steht für die Weiterverarbeitung als globale Variable zur Verfügung:



Datensatz der Unterscheidungsfunktionen gemäß Gleich. (15) bis (18) und (21) / (22)



Datensatz der Eigenschaftsklassen und der Vokalklassenzuordnung gemäß Tab. 1

Auf der Grundlage dieses Datensatzes können nun Schärfewerte berechnet werden, die im Zusammenhang mit den berechneten Vokalpräsenzen eine Aussage zur Vokalqualität ermöglichen sollen.

3.6 Simulation der Empfindungsgröße „Schärfe“

[Gliederung](#)

Eine Wahrnehmung ist ein Sinneskomplex, der sich aus Empfindungen zusammensetzt. Streng genommen handelt es sich bei der „Schärfe“ um eine Wahrnehmungsgröße.

In DIN 45692 [12] wird dazu folgende Aussage getroffen:

„Die Hörempfindung Schärfe beschreibt denjenigen Aspekt der Klangfarbenwahrnehmung, der mit der Frequenzverteilung der spektralen Hüllkurve von Schallen korreliert ist. Darüber hinaus können sich durch die spektrale Feinstruktur sowie durch Änderungen der zeitlichen Hüllkurve Klangfarbenunterschiede ergeben, die nicht durch die Schärfe erfasst werden.

Die Schärfe hat sich für die subjektive Beurteilung von Klangbildern bewährt. Mit zunehmender Schärfe werden Geräusche üblicherweise als aggressiver und häufig auch als lästiger empfunden.“

Die Schärfe S wird nach DIN 45692 [12] wie folgt berechnet:

$$S = k \cdot \frac{\int_{z=0}^{24 \text{ Bark}} N'(z) \cdot g(z) \cdot z \cdot dz}{\int_{z=0}^{24 \text{ Bark}} N'(z) \cdot dz} \text{ acum} \quad (25)$$

Darin ist k die Normierungskonstante zum Abgleich der berechneten Schärfe eines Referenzschalls auf $S = 1 \text{ acum}$ ($0,105 \leq k < 0,115$) und $g(z)$ die Gewichtsfunktion. Mit Gleich. (25) wird demzufolge ein gewichteter Tonheitsschwerpunkt berechnet. Bild 11 zeigt den Verlauf der Gewichtsfunktion.

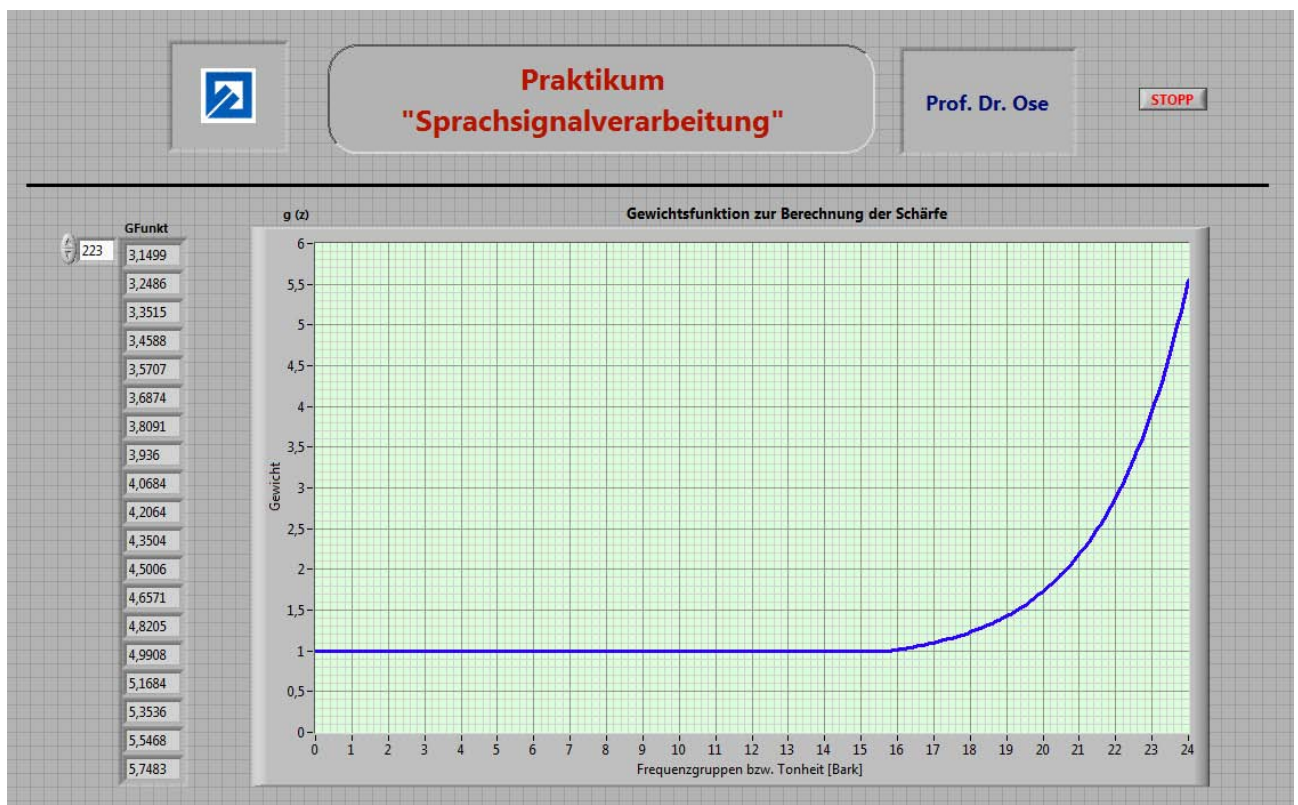


Bild 11: Gewichtsfunktion zur Berechnung der Schärfe nach [12]

Für den Referenzschall (Testsignal) gilt: $L_G = 60 \text{ dB}$ ($f = 1 \text{ kHz}$) $\Rightarrow S = 1 \text{ acum}$ [12]

Die Gewichtsfunktion kann nach [12] wie folgt berechnet werden:

$$\begin{aligned}
 g(z) &= 1 && \text{für } z \leq 15,8 \text{ Bark} \\
 g(z) &= 0,15 \cdot e^{0,42 \cdot (z/\text{Bark} - 15,8)} + 0,85 && \text{für } z > 15,8 \text{ Bark}
 \end{aligned}
 \tag{26}$$

Bild 12 zeigt einen ersten Ansatz zur Bestimmung der Schärfe nach Gleich. (25) und (26). Die Grafik im unteren Teil des Bildes 12 stellt den Verlauf der spezifischen Schärfe pro Frequenzgruppe dar.

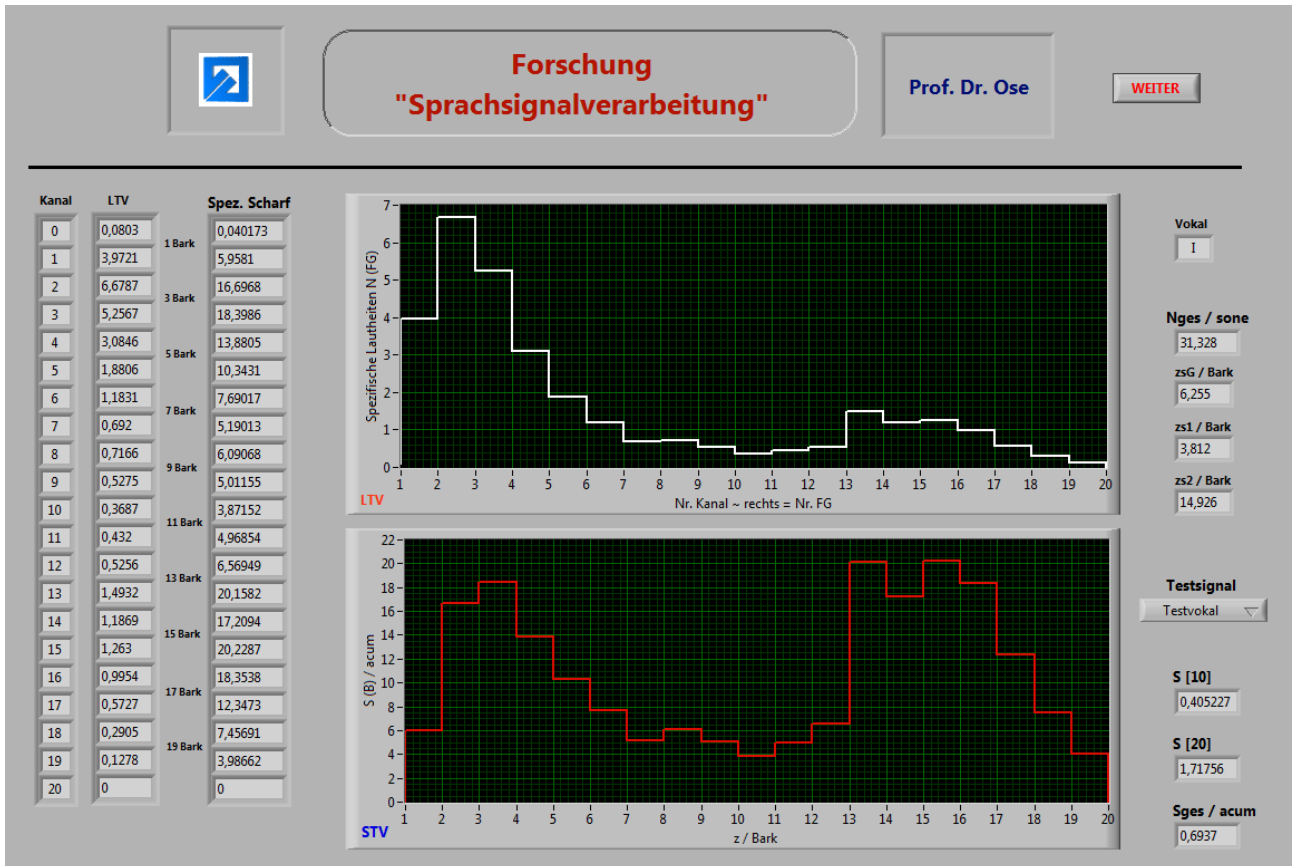


Bild 12: Vergleich zwischen spezifischer Lautheit und spezifischer Schärfe (Vokal | i: |)

Der resultierende Gesamtwert (hier: $S_{ges} \approx 0,7$ acum) ist zunächst wenig aussagefähig. Aus diesem Grund wurden die berechneten Schärfewerte aller Vokale eines Sprechers miteinander verglichen.

Tabelle 2: Schärfewerte der Vokale eines Testsprechers

Vokal	a	e	i	o	u	ä	ö	ü
$S_{ges} / acum$	0,86	0,91	1,01	0,63	0,52	0,99	0,82	0,81

Vokale sortiert	u	o	ü	ö	a	e	ä	i
$S_{ges} / acum$	0,52	0,63	0,81	0,82	0,86	0,91	0,99	1,01

Der Vergleich zeigt, dass die Schärfewerte der einzelnen Vokale nur unwesentlich voneinander abweichen. Es ist aber aus psychoakustischen Untersuchungen [4] bekannt, dass die Vokale einen deutlich voneinander verschiedenen Schärfeeindruck verursachen.

Eine Auswertung von psychoakustischen Untersuchungen (Hörtests mit 10 Probanden – Testmaterial: Vokale von 10 Sprechern – jeder Vokal dreimal in der Testfolge – zufällige Anordnung in der Folge) führt zu folgender mittleren Aussage (Skala: sehr stumpf bis sehr scharf):

Tabelle 3: Schärfewerte aus psychoakustischen Experimenten

Attribut	sehr stumpf	stumpf	mittel	scharf	sehr scharf
Skalenwert	0	0,5	1	1,5	2
Vokal	u	o	a ü/ö	ä e	i

Ein Vergleich der Tabellen 2 und 3 zeigt, dass die Tendenz der Skalenwerte abgebildet wird. Die in Tabelle 2 berechneten Zahlenwerte sind allerdings in der vorliegenden Form nicht akzeptabel.

Zur Überprüfung der zuverlässigen Arbeitsweise der programmierten Berechnungsvorschrift wurde der in Bild 13 dargestellte Test durchgeführt.

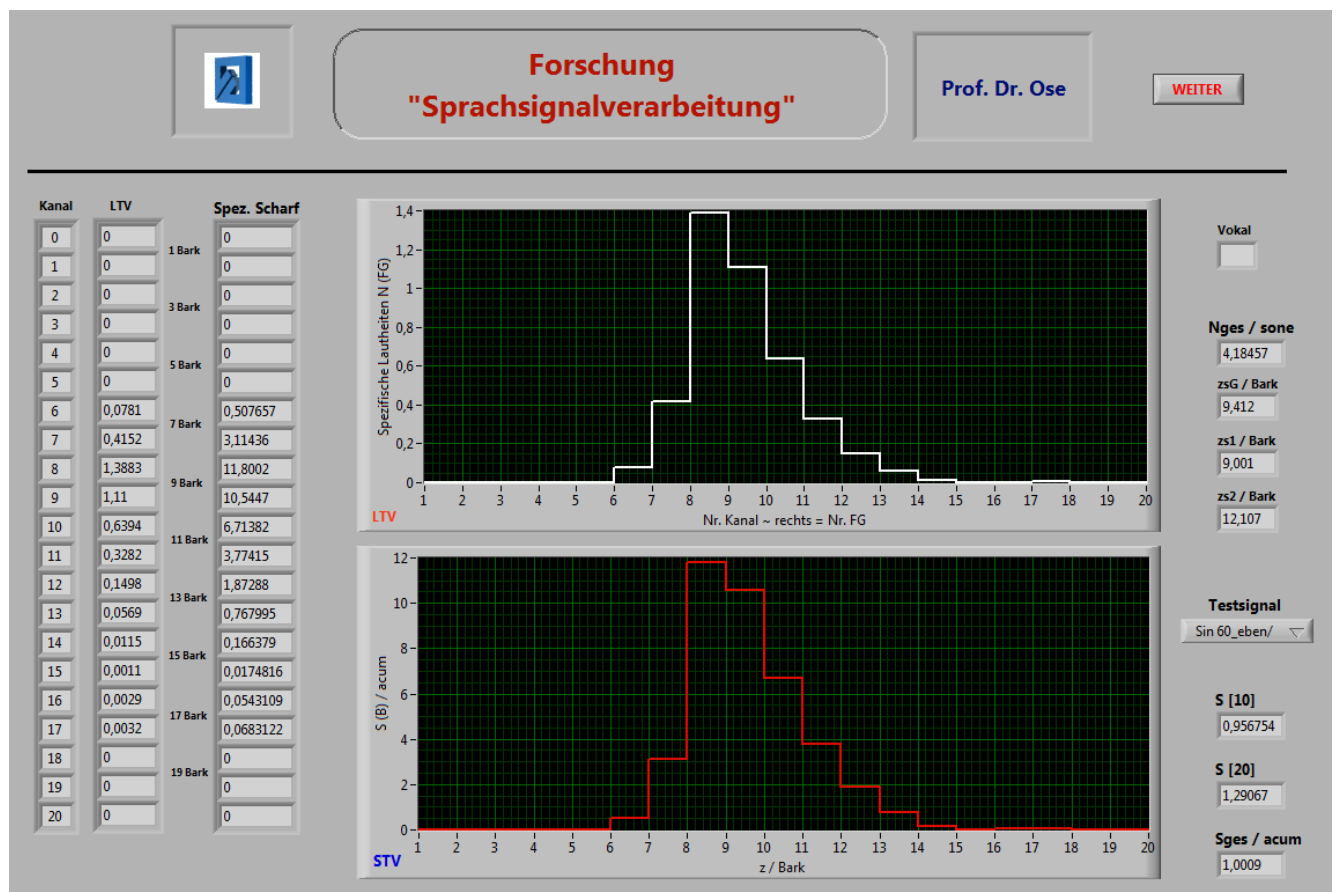


Bild 13: Verlauf der spezifischer Lautheit und der spezifischer Schärfe für das in [12] definierte Testsignal

Das in DIN 45692 [12] festgelegte Testsignal führt zu einer Schärfe mit $S_{ges} = 1 \text{ acum}$. Das Programm berechnet demzufolge die Schärfe zuverlässig.

In den weiteren Untersuchungen muss eine veränderte Berechnungsvorschrift entwickelt werden, die die Unterschiede in der vokalspezifischen Schärfe deutlicher abbildet.

4 Umstellung auf ein Analysesystem

[Gliederung](#)

Für die weiteren Untersuchungen (Ermittlung der Vokalqualität aller acht deutschen Langvokale) wird ein Analysesystem benötigt, das alle erforderlichen Verarbeitungsschritte des Experimentiersystems in möglichst automatisierter Form abarbeitet. Nur so ist eine statistische Absicherung (ca. 80 bis 100 Sprecher) der aus den Untersuchungsergebnissen abgeleiteten Schlussfolgerungen möglich.

Dazu sind zunächst die einzelnen Unterprogramme des Experimentiersystems zu optimieren (vgl. auch Kap. 5). Auf die Anzeige von Zwischenergebnissen sollte weitgehend verzichtet werden (Ausnahme: Festlegung des Analysezeitfensters). Bild 14 zeigt einen ersten Lösungsansatz zur Analyse einer Sprachprobe (hier: Männlicher Sprecher mit einer Grundfrequenz von 137 Hz – Vokal | i |).



Bild 14: Analyseergebnisse einer Testsprachprobe (Vokal | i: | aus „Uschi“)

Die Ergebnisse können als EXCEL-Liste und als Textdatei abgespeichert werden. Der Dateipfad wird automatisch generiert. Über eine Eingabeaufforderung ist dann die Festlegung des Dateinamens möglich.

Die Daten im Bild 14 sind noch unvollständig. Es fehlt noch der Datensatz zur Beurteilung der Vokalqualität. In diesem Zusammenhang wäre auch eine grafische Darstellung in Form eines Vergleichs zwischen Testsprachprobe und Referenzsprachprobe wünschenswert.

5 Derzeitiger Leistungsstand

[Gliederung](#)

5.1 Aufnahme von Sprachproben (vgl. Kap. 2)

Das entwickelte Aufnahmesystem hat sich bewährt. Die Aussteuerungsanzeige funktioniert hervorragend und stellt ein wichtiges Hilfsmittel für den Testsprecher dar. Die manuelle Eingabe des gemessenen Schallpegels ist gewöhnungsbedürftig (aber akzeptabel).

Das Abspeichern der Datensätze ist infolge der automatischen Generierung der Dateinamen problemlos möglich. Es wurden bisher Sprachproben von ca. 30 Sprechern aufgenommen. Hier gibt es weiteren Handlungsbedarf (Ziel: 80 bis 100 Sprecher – vgl. auch Anlage 5).

Die Sprachproben aller Testsprecher befinden sich im Ordner: c:\Sprache

5.2 Hauptprogramm (vgl. Abschn. 3.1)

Das Hauptprogramm ist erweiterungsfähig und hat sich in der vorliegenden Form im Sinne eines Experimentiersystems bewährt. Eine Umstellung auf ein Analysesystem (vgl. auch Kap. 4) sollte problemlos möglich sein. Dazu ist es erforderlich, dass alle notwendigen Unterprogramme in der entsprechenden Reihenfolge abgearbeitet werden. Alle ermittelten Datensätze werden über globale Variable an das Unterprogramm „Auswertung“ übergeben.

5.3 Bestimmung von Frequenzgruppenpegeln (vgl. Abschn. 3.2)

Dieses Unterprogramm hat die Aufgabe, die physikalischen Größen zur Berechnung eines psychisch relevanten Merkmalsatzes bereitzustellen. Die Ergebnisse dieses Unterprogramms bilden demzufolge die Grundlage für die Genauigkeit / Aussagefähigkeit der nachfolgend abzuleitenden Analysedaten.

- Festlegung des Analysezeitfensters:

Die hier gefundene Lösung ist äußerst leistungsfähig. Das Analysezeitfenster kann mit der eingestellten Fensterbreite über der vorgeschrittenen Sprachprobe verschoben werden. Dabei sorgt die verwendete AKF für einen definierten Startzeitpunkt und eine reproduzierbare Navigation des Fensters längs der Sprachprobe.

- Darstellung des Grundfrequenzverlaufs:

Der Verlauf der Grundfrequenz liefert dem Bearbeiter wichtige Zusatzinformationen zur Positionierung des Analysezeitfensters und zu quasistationären Bereichen in der vorgeschrittenen Sprachprobe.

- Berechnung und Darstellung des Linienspektrums:

Das Linienspektrum wird für die nachfolgenden Analysen nicht benötigt. Es stellt aber für den Bearbeiter ein wichtiges Hilfsmittel bei der subjektiven Beurteilung der Ausprägung von Formanten (Maxima in der Hüllkurve des Amplitudenspektrums) bereit.

- Berechnung und Darstellung der Frequenzgruppenpegel:

Die Frequenzgruppenpegel bilden die Grundlage zur nachfolgenden Berechnung von Erregungspegeln und Kernlautheiten. Das vorliegende Unterprogramm ermittelt die Frequenzgruppenpegel exakt nach der Theorie von ZWICKER [5]. Die Lautheits-Tonheits-Verteilung wird aber mit genäherten Frequenzgruppen (Zusammenfassung von Terzbändern und Zuordnung zu den tatsächlichen Frequenzgruppen – siehe [6], [10] und [11]) berechnet. Wie man in Anlage 1 (Tab. A_1.5) leicht sieht, führt diese Maßnahme oberhalb von ca. 10 Bark zu massiven Abweichungen. Das kann in dieser Form nicht akzeptiert werden !

In [6] und [10] werden Tabellen (vgl. auch Anlage 1) zur Berechnung der Lautheits-Tonheits-Verteilung angegeben, die sich mit unterschiedlichen Werten sogar teilweise widersprechen. Sie sind auf die

Anwendung genäherter Frequenzgruppen ausgerichtet und gelten auch nur für diesen Fall. Andere Tabellen (das sind Ergebnisse umfangreicher und jahrelanger psychoakustischer Untersuchungen – siehe [5]) gibt es leider nicht.

Für die weitere Bearbeitung des Themas wäre folgender Kompromiss denkbar:

- a) Die Parallelfilterbank wird (eventuell zwischen Terzbänder und Frequenzgruppen umschaltbar) auf Terzbänder (genäherte Frequenzgruppen) umgestellt.
- b) Die nachfolgende Berechnung von Erregungspegeln und Kernlautheiten sowie die Berechnung der Lautheits-Tonheits-Verteilung werden mit den genäherten Frequenzgruppen durchgeführt. Man erhält 240 Stützwerte der spezifischen Lautheit in 20 genäherten Frequenzgruppen.
- c) Die Stützwerte werden so zusammengefasst, dass man die mittleren spezifischen Lautheiten in den 24 (tatsächlichen) Frequenzgruppen erhält.

5.4 Berechnung von Erregungspegeln und Kernlautheiten (vgl. Abschn. 3.3)

Für die Kernlautheiten existieren drei Berechnungsvorschriften [vgl. Gleich. (9), (10) und (11)], die je nach Signalkonfiguration zu etwas voneinander abweichenden Ergebnissen führen. Nach der Umsetzung der in Abschn. 5.3 vorgeschlagenen Maßnahme muss eine Entscheidung über eine einheitliche Berechnung der Kernlautheit getroffen werden. Das ist relativ einfach möglich, da bereits zwei Berechnungsvorschriften (umschaltbar) einprogrammiert wurden.

Ansonsten läuft das Unterprogramm sehr zuverlässig. Bei der Umstellung auf ein Analysesystem kann auf die grafischen Anzeigen verzichtet werden, wenn der Nachweis erbracht ist, dass definierte Testsignale zu einem reproduzierbaren Ergebnis führen.

5.5 Berechnung der Lautheits-Tonheits-Verteilung (vgl. Abschn. 3.4)

Das Unterprogramm selbst erfüllt zuverlässig seine Funktion. Die Ursachen für Abweichungen von den in der DIN 45631 [10] vorgeschlagenen Testsignalen sind bekannt (siehe Abschn. 5.3 und 5.4).

5.6 Vokalklassifikation (vgl. Abschn. 3.5)

Das Unterprogramm zur Berechnung der Vokalpräsenzen und zur Vokalklassifikation reagiert auf die in Abschn. 5.3 bis 5.5 diskutierten Unregelmäßigkeiten wenig sensibel. Das ist erstaunlich und lässt die Hoffnung zu, dass die berechneten Unterscheidungsfunktionen infolge der Summenbildung gar nicht so sehr von der Feinstruktur des Merkmalsatzes abhängig sind.

5.7 Simulation der Empfindungsgröße „Schärfe“ (vgl. Abschn. 3.6)

Die Ergebnisse zur Berechnung der Schärfe sind unbefriedigend. Wenn man jedoch die Tatsache berücksichtigt, dass es sich bei der Schärfe nach DIN 45692 [12] ohne die Normierungskonstante und ohne die Gewichtsfunktion [vgl. Gleich. (25)] lediglich um den aus der Vokalklassifikation bekannten Tonheitsschwerpunkt handelt, können erste Schlussfolgerungen zu diesen Ergebnissen abgeleitet werden:

- a) Die Normierungskonstante „drückt“ den Schärfewert in Richtung „1“ und verwischt so die Unterschiede zwischen Vokalen mit unterschiedlicher Signalkonfiguration.
- b) Die Gewichtsfunktion wird erst in demjenigen Tonheitsbereich wirksam (ab ca. 16 Bark), in dem die Abweichungen zwischen den Frequenzgruppen und den Terzbändern immer gravierender werden. Die im Abschn. 5.3 vorgeschlagene Maßnahme könnte dazu beitragen, die Wirksamkeit der Gewichtsfunktion zu erhöhen. Eine Trennung der Tonheitsskala nach Vorbild der Vokalklassifikation bei $z_a = 10,5$ Bark in zwei Teilbereiche und eine getrennte Schärferechnung in beiden Bereichen könnte zu einem erfolversprechenden Lösungsansatz führen.

Anlage 1: Tabellen[Gliederung](#)

Tabelle A_1.1: Grenzwerte der Frequenzgruppen nach [5]

Kanal-Nr.	z_{gu} / Bark	z_{go} / Bark	f_{gu} / Hz	f_{go} / Hz
		Nr. FG		
0	0	1	20	100
1	1	2	100	200
2	2	3	200	300
3	3	4	300	400
4	4	5	400	510
5	5	6	510	630
6	6	7	630	770
7	7	8	770	920
8	8	9	920	1080
9	9	10	1080	1270
10	10	11	1270	1480
11	11	12	1480	1720
12	12	13	1720	2000
13	13	14	2000	2320
14	14	15	2320	2700
15	15	16	2700	3150
16	16	17	3150	3700
17	17	18	3700	4400
18	18	19	4400	5300
19	19	20	5300	6400
20	20	21	6400	7700
21	21	22	7700	9500
22	22	23	9500	12000
23	23	24	12000	15500

Tabelle A_1.2: Dämpfungsmaß α_0 des Gehörs, Pegeldifferenz ΔL_{ED} zwischen ebenem und diffusem Schallfeld und Erregungspegel L_{EHS} an der Hörschwelle bei den Mittenfrequenzen f_m von 24 Frequenzgruppen [6]

Nr. FG	f_{g0} / Hz	f_m / Hz	α_0 / dB	ΔL_{ED} / dB	L_{EHS} / dB
1	100	50	0	0	42
2	200	150	0	0,1	18,5
3	300	250	0	0,6	11,5
4	400	350	0	1,0	8,3
5	510	450	0	1,4	6,7
6	630	570	0	2,0	5,5
7	770	700	0	2,5	4,8
8	920	840	0	2,8	4,3
9	1080	1000	0	3,0	4,0
10	1270	1170	0	2,3	4,0
11	1480	1370	- 0,2	1,3	4,0
12	1720	1600	- 0,5	0	4,0
13	2000	1850	- 1,2	- 0,9	4,0
14	2320	2150	- 2,1	- 1,6	4,0
15	2700	2500	- 3,2	- 2,0	4,0
16	3150	2900	- 4,6	- 1,9	4,0
17	3700	3400	- 5,5	- 1,6	4,0
18	4400	4000	- 5,6	- 1,0	4,0
19	5300	4800	- 4,3	0,2	4,0
20	6400	5800	- 2,5	2,0	4,0
21	7700	7000	- 0,1	3,4	4,0
22	9500	8500	2,8	4,1	4,0
23	12000	10500	6,4	4,2	4,0
24	15500	13500	20,0	3,5	4,0

Anmerkung:

Diese Tabelle wird in der DIN 45631 [10] nicht mehr angegeben. Die Tabellenwerte stimmen nicht mit den Werten der Tabelle A_1.3 überein. Dort werden die Daten für Terzbänder angegeben, während die Werte der Tabelle A_1.2 für Frequenzgruppen gültig sind.

Tabelle A_1.3: Dämpfungsmaß α_0 des Gehörs, Pegeldifferenz ΔL_{ED} zwischen ebenem und diffusem Schallfeld, Erregungspegel L_{EHS} an der Hörschwelle und Korrekturgröße ΔL_T bei den Mittenfrequenzen f_m von 20 genäherten Frequenzgruppen bzw. Terzen [6] [10]

Terz \approx FG	f_m / Hz	f_{go} / Hz	z_{go} / Bark	α_0 / dB	ΔL_{ED} / dB	L_{EHS} / dB	ΔL_T / dB
1	63	90	0,9	0	0	30,0	-0,25
2	125	180	1,8	0	0	18,0	-0,6
3	224	280	2,8	0	0,5	12,0	-0,8
4	315	355	3,5	0	0,9	8,0	-0,8
5	400	450	4,4	0	1,2	7,0	-0,5
6	500	560	5,4	0	1,6	6,0	0
7	630	710	6,6	0	2,3	5,0	0,5
8	800	900	7,9	0	2,8	4,0	1,1
9	1000	1120	9,2	0	3,0	3,0	1,5
10	1250	1400	10,6	0	2,0	3,0	1,7
11	1600	1800	12,3	-0,5	0	3,0	1,8
12	2000	2240	13,8	-1,6	-1,4	3,0	1,8
13	2500	2800	15,2	-3,2	-2,0	3,0	1,7
14	3150	3550	16,7	-5,4	-1,9	3,0	1,6
15	4000	4500	18,1	-5,6	-1,0	3,0	1,4
16	5000	5600	19,3	-4,0	0,5	3,0	1,2
17	6300	7100	20,6	-1,5	3,0	3,0	0,8
18	8000	9000	21,8	2,0	4,0	3,0	0,5
19	10000	11200	22,7	5,0	4,3	3,0	0
20	12500	14000	23,6	12,0	4,0	3,0	-0,5

Anmerkung:

Die Werte in der gelb gekennzeichneten Spalte dieser Tabelle wurden in der DIN 45631 [10] gegenüber den Tabellenwerten in [6] verändert.

Tabelle A_1.4: Flankensteilheit – dN'/dz in sone_G/Bark für 16 Wertebereiche von N' in Abhängigkeit von der durchlaufenden Frequenzgruppe oder Terz [6]

WB	Untergrenz	Frequenzgruppe oder Terz							
Nr.	sone/Bark	2	3	4	5	6	7	8	> 8
1	21,5	13,0	8,2	6,3	5,5	5,5	5,5	5,5	5,5
2	18,0	9,0	7,5	6	5,1	4,5	4,5	4,5	4,5
3	15,1	7,8	6,7	5,6	4,9	4,4	3,9	3,9	3,9
4	11,5	6,2	5,4	4,6	4,0	3,5	3,2	3,2	3,2
5	9,0	4,5	3,8	3,6	3,2	2,9	2,7	2,7	2,7
6	6,1	3,7	3,0	2,8	2,35	2,2	2,2	2,2	2,2
7	4,4	2,9	2,3	2,1	1,9	1,8	1,7	1,7	1,7
8	3,1	2,4	1,7	1,5	1,35	1,3	1,3	1,3	1,3
9	2,13	1,95	1,45	1,3	1,15	1,1	1,1	1,1	1,1
10	1,36	1,5	1,2	0,94	0,86	0,82	0,82	0,82	0,82
11	0,82	0,72	0,67	0,64	0,63	0,62	0,62	0,62	0,62
12	0,42	0,59	0,53	0,51	0,5	0,42	0,42	0,42	0,42
13	0,3	0,4	0,33	0,26	0,24	0,22	0,22	0,22	0,22
14	0,22	0,27	0,21	0,2	0,18	0,17	0,17	0,17	0,17
15	0,15	0,16	0,15	0,14	0,12	0,11	0,11	0,11	0,11
16	0,1	0,12	0,11	0,1	0,08	0,08	0,08	0,08	0,08
17	0,035	0,09	0,08	0,07	0,06	0,06	0,06	0,06	0,05
18	0	0,06	0,05	0,03	0,02	0,02	0,02	0,02	0,02

Anmerkung:

Die Werte in dieser Tabelle wurden in der DIN 45631 [10] gegenüber den Tabellenwerten in [6] verändert und um zwei Zeilen erweitert.

Tabelle A_1.5: Vergleich der Grenzwerte von Terzbändern und Frequenzgruppen

Frequenzgruppen				Terzbänder *			
z_{gu} / Bark	z_{go} / Bark	f_{gu} / Hz	f_{go} / Hz	z_{gu} / Bark	z_{go} / Bark	f_{gu} / Hz	f_{go} / Hz
	Nr. FG				ZUP		
0	1	20	100	0	0,9	20	90
1	2	100	200	0,9	1,8	90	180
2	3	200	300	1,8	2,8	180	280
3	4	300	400	2,8	3,5	280	355
4	5	400	510	3,5	4,4	355	450
5	6	510	630	4,4	5,4	450	560
6	7	630	770	5,4	6,6	560	710
7	8	770	920	6,6	7,9	710	900
8	9	920	1080	7,9	9,2	900	1120
9	10	1080	1270	9,2	10,6	1120	1400
10	11	1270	1480	10,6	12,3	1400	1800
11	12	1480	1720	12,3	13,8	1800	2240
12	13	1720	2000	13,8	15,2	2240	2800
13	14	2000	2320	15,2	16,7	2800	3550
14	15	2320	2700	16,7	18,1	3550	4500
15	16	2700	3150	18,1	19,3	4500	5600
16	17	3150	3700	19,3	20,6	5600	7100
17	18	3700	4400	20,6	21,8	7100	9000
18	19	4400	5300	21,8	22,7	9000	11200
19	20	5300	6400	22,7	23,6	11200	14000
20	21	6400	7700	23,6	24	14000	15500
21	22	7700	9500				
22	23	9500	12000				
23	24	12000	15500				

Anlage 2: Programmablaufplan zur Berechnung der Lautheit nach [6]

[Gliederung](#)

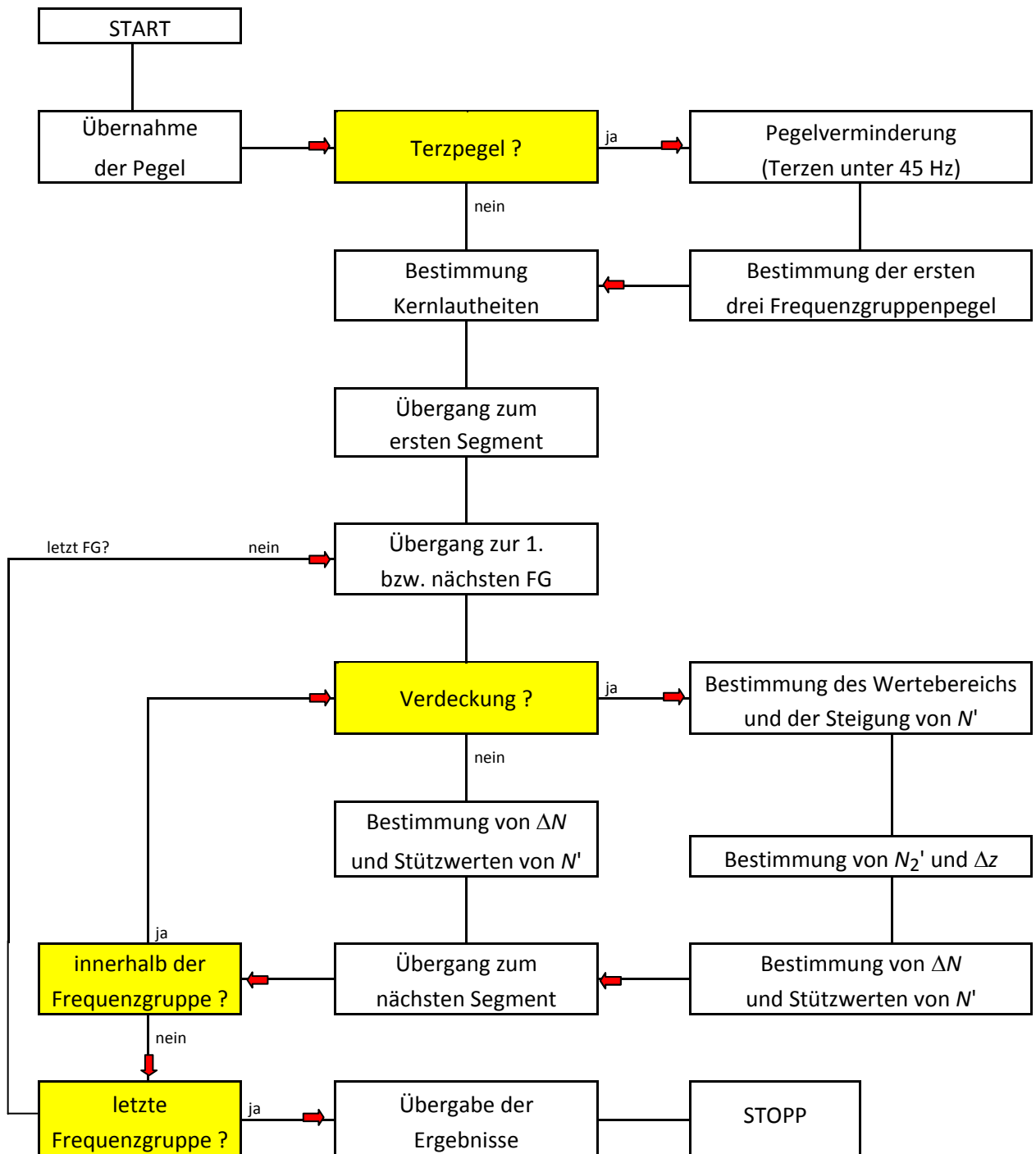


Bild A_2.1: Programmablaufplan zur Lautheitsberechnung

Anlage 3: Formelzeichenverzeichnis und Abkürzungen[Gliederung](#)

Tabelle A_3.1: Formelzeichen (nach Bezeichnung alphabetisch sortiert)

Bezeichnung	Formelzeichen	Einheit	Bemerkung
Abtastfrequenz	f_a	Hz	auch: Samples / s
Autokorrelations-fkt.	AKF		
Bandbreite	B	Hz	
Bezugsschallpegel	L_{Bezug}	dB	
Cursorposition	X		
Dämpfung	A	dB	
Dämpfung FG	A_G	dB	
Dämpfung Stützwert	A_{St}	dB	
Dämpfungsmaß	a_0	dB	
Eigenschaftsgruppe	EG		
Eigenschaftsklasse	e_k		
Erregungspegel	L_E	dB	
Erregungspegel HS	L_{EHS}	dB	
Fensterbreite	t_F	s	auch: F in Samples
FG-Pegel	L_{St}	dB	
Frequenz	f	Hz	
Frequenz Harmon.	f_{Hn}	Hz	
Frequenzabstand	Δf	Hz	Stützwerte
Frequenzgruppe	FG		
Grenzfrequenz	f_g	Hz	
Grenzscherpunkt	z_{SG}	Bark	
Grenztonheit	z_g	Bark	
Grundfrequenz	f_0	Hz	
Hörschwelle	HS		
Kernlautheit	N'_{Kern}	sone _G / Bark	
Konstante	K		
Korrekturpegel Terz	ΔL_T	dB	
Lautheit	N	sone	
Merkmalvektor	\vec{N}	sone _G / Bark	
Mittenfrequenz	f_m	Hz	
Nr. Harmonische	n_H		
Pegeldifferenz	ΔL_{ED}	dB	eben – diffus
Ruhehörschwelle	RHS		
Schallpegel	L	dB	
Schallpegel Harmon.	L_{Hn}	dB	
Schallpegel Stützwert	L_{St}	dB	
Schärfe	S	acum	
Schwellenfaktor	s		
spezifische Lautheit	$N'(z)$	sone _G / Bark	

Tonheit	z	Bark	
Tonheitsschwerpunkt	z_s	Bark	
Unterscheidungsfkt.	$d(\vec{N})$		
Vokalklasse	$[k]$		
Wertebereich	WB		
Zähler	n		
Zähler Filterkanäle	j		
Zähler Stützwerte	i		

Anlage 4: Testsignale nach DIN 45631:**gültig für freies (ebenes) Schallfeld !!!**

Tabelle A_4.1: Daten der genäherten Frequenzgruppen zur Lautheitsberechnung

[Gliederung](#)

Nr. Terz	f_m (Terz) / Hz	Nr. der genäherten FG	f_{g0} / Hz der genäherten FG	Terzpegel in dB	Lautheit in sone
T1	25	1	90		
T2	31,5				
T3	40				
T4	50				
T5	63				
T6	80				
T7	100	2	180		
T8	125				
T9	160				
T10	200	3	280		
T11	250				
T12	315	4	355		
T13	400	5	450		
T14	500	6	560	10	
T15	630	7	710	30	
T16	800	8	900	50	
T17	1000	9	1120	70	8
T18	1250	10	1400		
T19	1600	11	1800		
T20	2000	12	2240		
T21	2500	13	2800		
T22	3150	14	3550		
T23	4000	15	4500		
T24	5000	16	5600		
T25	6300	17	7100		
T26	8000	18	9000		
T27	10000	19	11200		
T28	12500	20	14000		

Alle anderen Pegelwerte = 0 (Eingabewerte = - 60 dB). Die um jeweils 20 dB abgesenkten Pegel unterhalb des Durchlassbereiches des Filters mit der Mittenfrequenz $f_m = 1$ kHz bilden die reale Filtercharakteristik nach.

Der berechnete Lautstärkepegel in phon entspricht bei $f = 1$ kHz dem Schallpegel bei $f = 1$ kHz in dB.

Weitere Testsignale: (alles bei $f = 1$ kHz)

40 dB (also: 20; 40)

$N_{ges} = 1$ sone

(Definition nach Zwicker)

50 dB (also: 10; 30; 50)

$N_{ges} = 2$ sone

60 dB (also: 20; 40; 60)

$N_{ges} = 4$ sone

80 dB (also: 20; 40; 60; 80)

$N_{ges} = 16$ sone

90 dB (also: 10; 30; 50; 70; 90)

$N_{ges} = 32$ sone

100 dB (also: 20; 40; 60; 80; 100)

$N_{ges} = 64$ sone

8 sone für 70 dB
(siehe oben)

A 4.5: Testpegel 80 dB ($f = 1$ kHz)

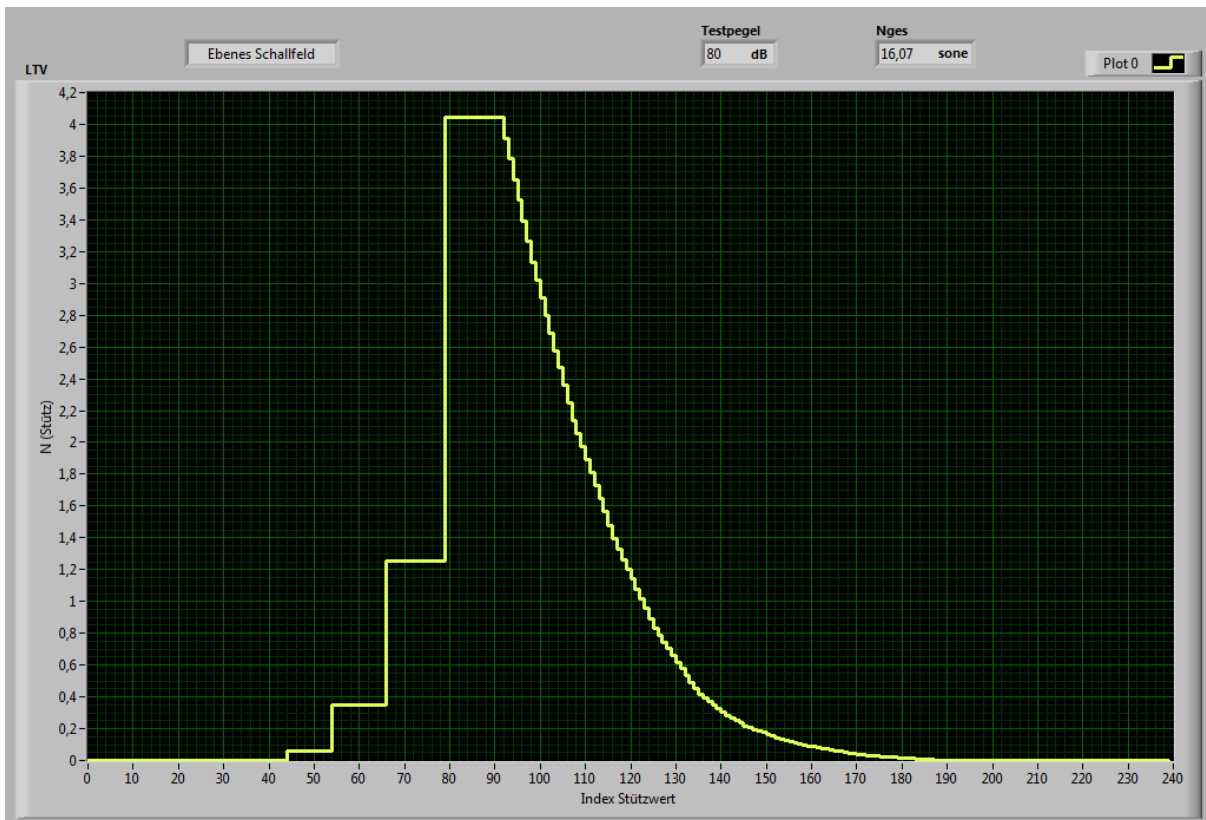


Bild A_4.5.1: Verlauf der Stützwerte bei einem Testpegel von 80 dB ($f = 1$ kHz)

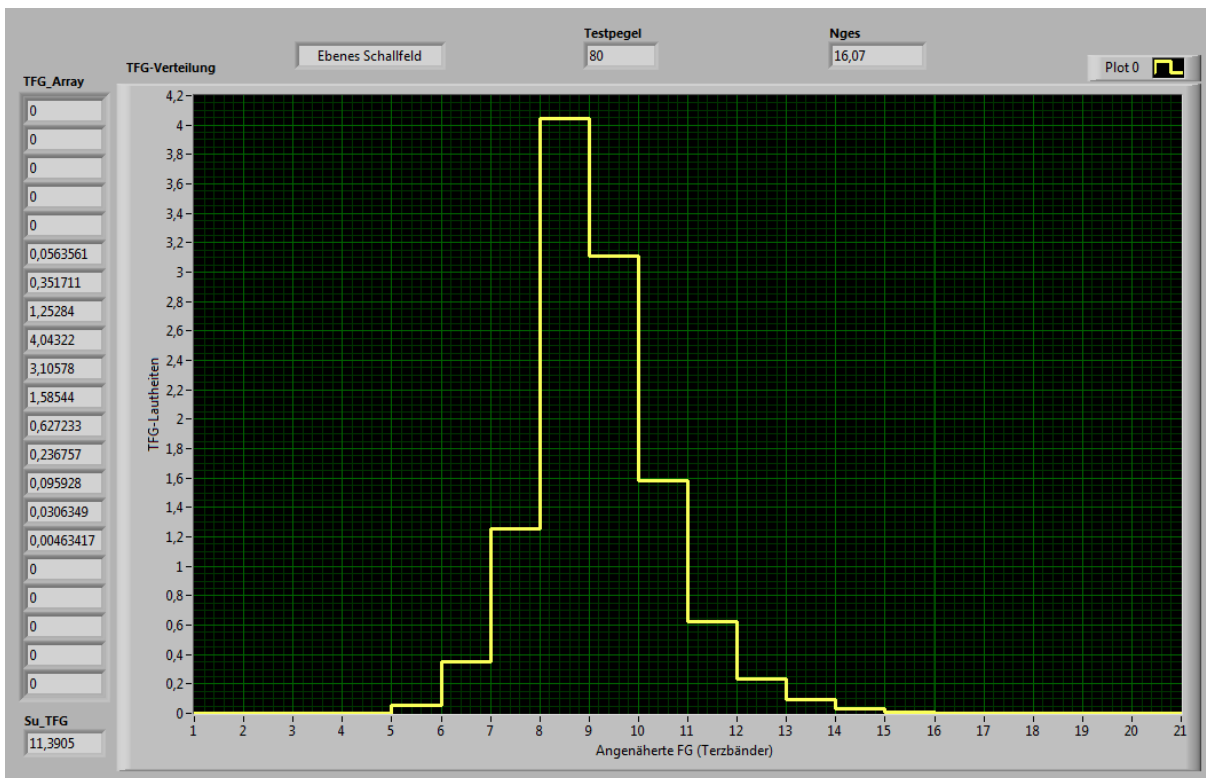


Bild A_4.5.2: Zusammenfassung der Stützwerte zu Terzbändern bei einem Testpegel von 80 dB ($f = 1$ kHz)

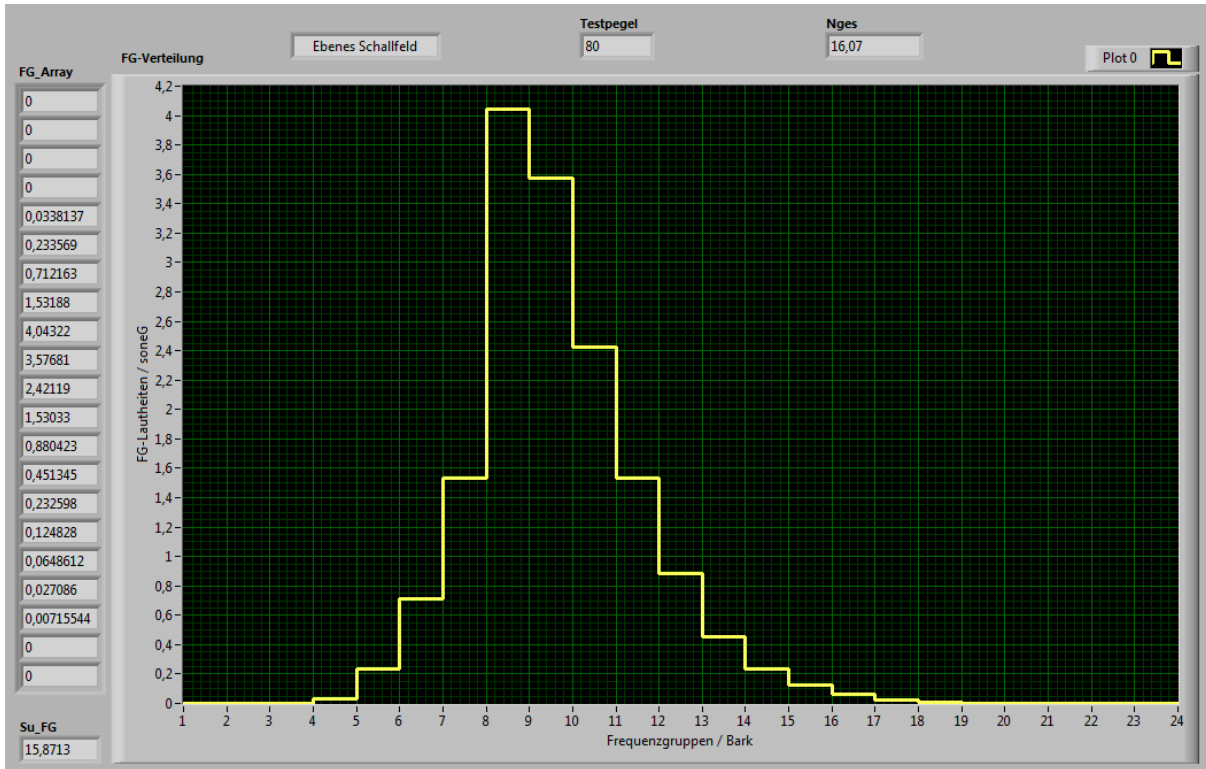


Bild A_4.5.3: Zusammenfassung der Stützwerte zu Frequenzgruppen bei einem Testpegel von 80 dB ($f = 1$ kHz)

A 4.6: Testpegel 90 dB ($f = 1$ kHz)

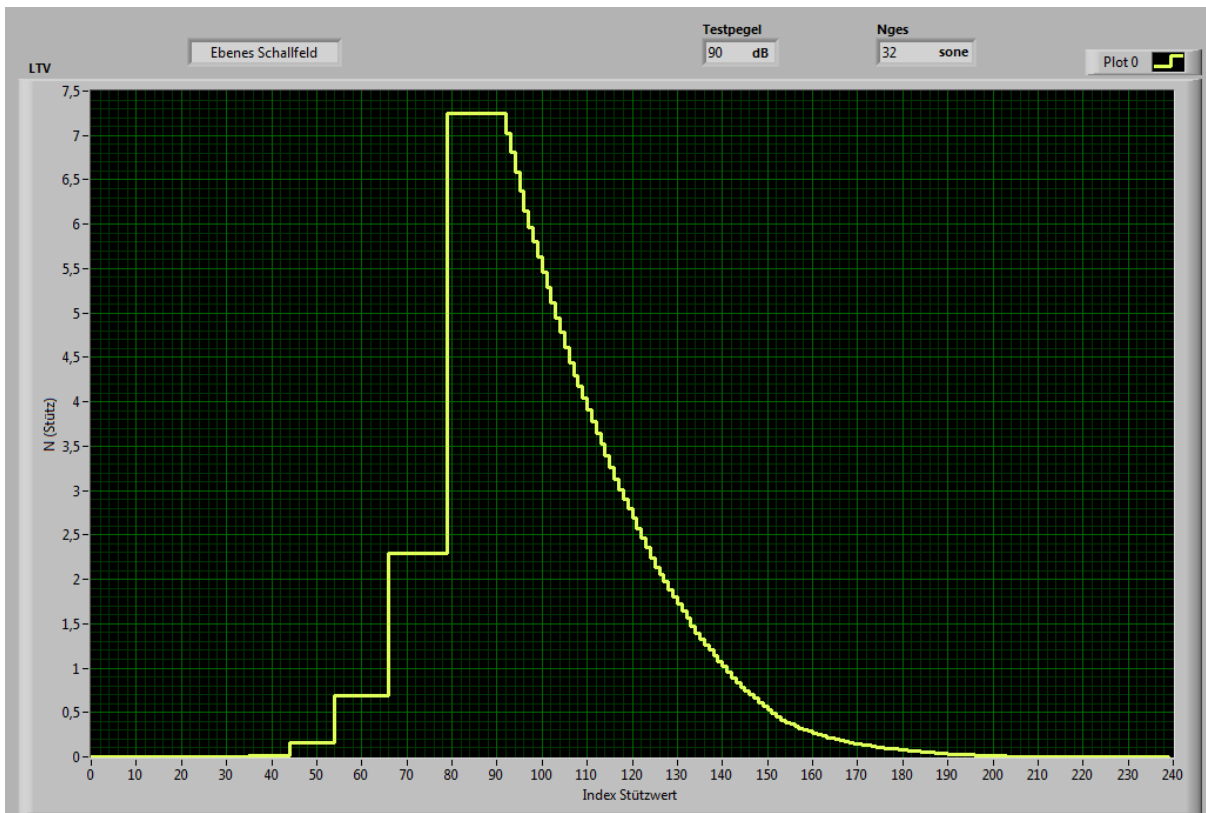


Bild A_4.6.1: Verlauf der Stützwerte bei einem Testpegel von 90 dB ($f = 1$ kHz)

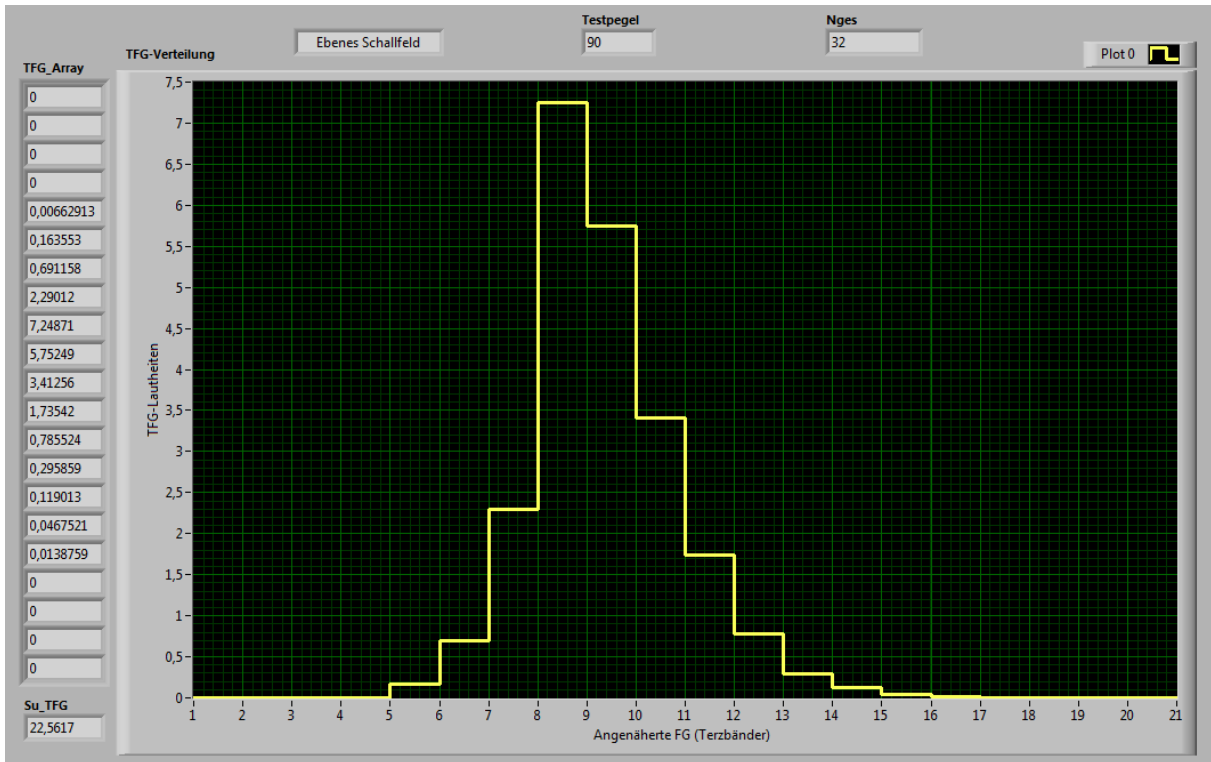


Bild A_4.6.2: Zusammenfassung der Stützwerte zu Terzbändern bei einem Testpegel von 90 dB ($f = 1$ kHz)

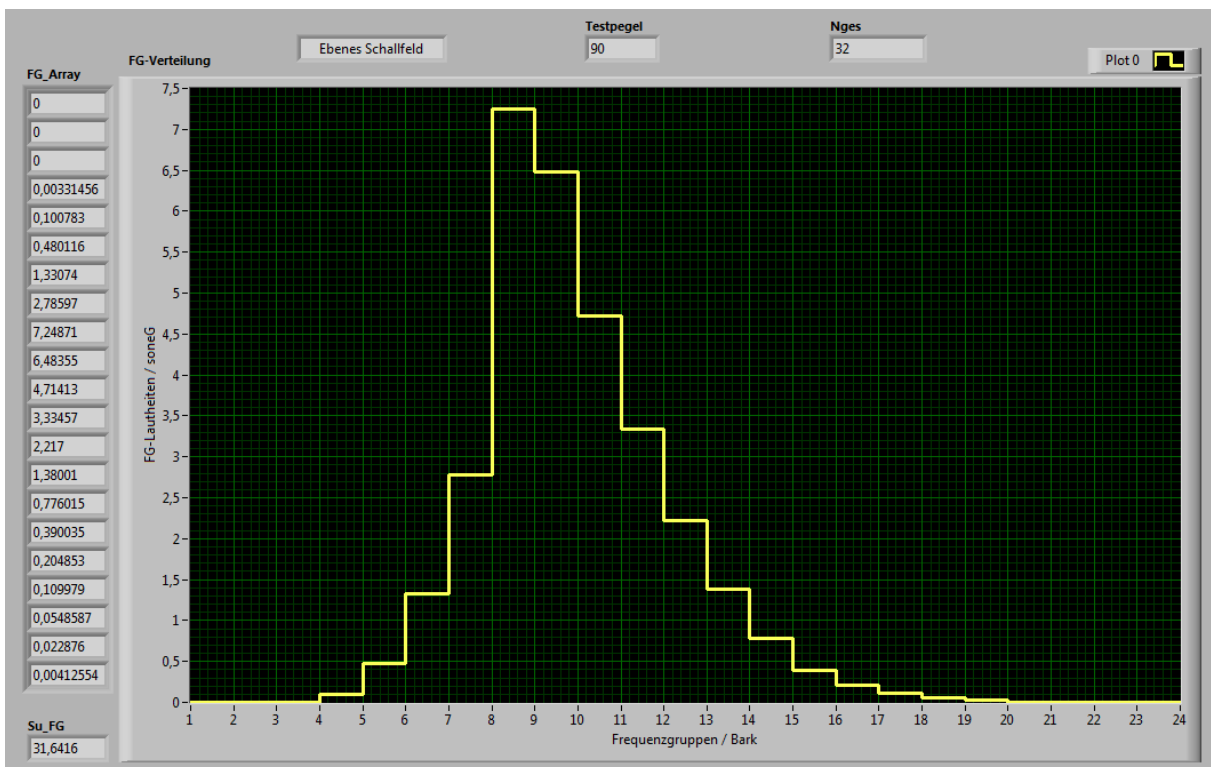


Bild A_4.6.3: Zusammenfassung der Stützwerte zu Frequenzgruppen bei einem Testpegel von 90 dB ($f = 1$ kHz)

Anlage 5: Arbeitspläne[Gliederung](#)**A 5.1: Arbeitsplan zum SS 2012**

März/April 2012: Entwicklung eines LabVIEW-Programms zur Berechnung von Lautheiten aus Frequenzgruppenpegeln

Mai 2012: Entwicklung eines LabVIEW-Programms zur Berechnung von Klangattributen aus Lautheits-Tonheits-Verteilungen

Juni/Juli 2012: Qualitätsbeurteilung auf der Basis von Klangattributen

Begründung für den Umfang: Programmtechnische Umsetzung von komplexen Verarbeitungsmechanismen des humanen Perzeptionssystems; die Ableitung von Schlussfolgerungen ist nur mit umfangreichen Hörtests möglich.

A 5.2: Präzisiertes Arbeitsplan zum SS 2012

Nr.	Aufgabe	Bearbeitungsstand
1	Aufnahme von Testsprachproben (Ziel: 80 bis 100 Sprecher / 8 Vokale)	bis 6 / 2012
	• Entwicklung eines Programms (Aussteuerungsanzeige/Pegelmessung)	fertig
	• automatisierte Aufnahme und Abspeicherung	fertig
	• Kalibrierungs- und Messvorschrift (Vorverstärker / Mikro)	fertig
2	Programm zum Laden und Schneiden einer Sprachprobe	am 1. 4. fertig
	• effektives Schneiden der wav-Dateien in automatisierter Form	fertig
	• automatische Auswahl eines optimalen Analyse-Zeitfensters (t_0 und t_F)	fertig
	• im Zeitfenster: eine FFT und eine Kanal-Analyse (Basis: FG)	fertig
	• Darstellung: Linienspektrum; Erregungspegel; FG-Pegel	fertig
3	Programm zur Berechnung der Lautheits-Tonheits-Verteilung	am 1. 7. fertig
	• Kernlautheiten und Flankenlautheiten	Ende April
	• Unterscheidungsfunktionen nach der Vokal-Theorie	Ende Mai
	• Vergleich mit DIN 45 631 (Suche nach Optimum)	Ende Juni
4	Programm zur Bestimmung der Vokalpräsenz	am 31. 7. fertig
5	Ausgabe des Analysedatensatzes in eine .xls-Datei und eine .doc-Datei	in Bearbeitung

A 5.3: Übertrag in den Arbeitsplan zum WS 2012 (muss im Oktober präzisiert werden)

9 + 10 / 2012: Entwicklung eines LabVIEW-Programms zur Berechnung ausgewählter Klangattribute (insbesondere Schärfe) aus Lautheits-Tonheits-Verteilungen

11 / 2012: Psychoakustische Tests zur Qualitätsbeurteilung vokalischer Realisierungen

12 / 2012: Programmtechnische Umsetzung der ermittelten Qualitätsmaße

01 / 2013: Entwicklung von Vorschlägen zur Synthese qualitativ hochwertiger vokalischer Klänge für einen eventuellen Einsatz in Elektromobilen

Literaturverzeichnis[Gliederung](#)

- [1] *Ose, R.*: Entwicklung und technische Realisierung eines interaktiven Sprachtrainers für Hör- und Sprechgeschädigte. Forschungsbericht. NWK (AgiP): Wolfenbüttel, 2001
- [2] *Ose, R.*: Analyse und Qualitätsbeurteilung von Sprachproben. Forschungsbericht. BMBF (AiF): Wolfenbüttel, 2005
- [3] *Ose, R.; Ilchmann, R.*: Handbuch COGIST 2000. Wolfenbüttel, 2005
<https://www.ostfalia.de/cms/de/pws/ose/forschung.htm>
- [4] *Ose, R.*: Extraktion von Merkmalen zur Vokalklassifikation. Dissertation. TU Dresden, 1984
- [5] *Zwicker, E.; Feldtkeller, R.*: Das Ohr als Nachrichtenempfänger. – 2. Auflage. – Stuttgart: S. Hirzel Verlag, 1967
- [6] *Paulus, E.; Zwicker, E.*: Programme zur automatischen Bestimmung der Lautheit aus Terzpegeln oder Frequenzgruppenpegeln. *Acustica* 27 (1972) 5
- [7] *National Instruments*: LabVIEW-Hilfe (8.6). FFT-Leistungsspektrum (VI).
- [8] *Ose, R.*: Modell zur Perzeption isolierter deutscher Langvokale. *Nachrichtentechnik / Elektronik* 37 (1987) 1. – Berlin: Verlag Technik, 1987
- [9] *Tscheschner, W.; Ose, R.*: Verfahren zur Vokaldiskrimination und -klassifikation. Patentschrift DD 243 791 A1. – Berlin: Amt für Erfindungs- und Patentwesen, 1987
- [10] *DIN 45631*: Berechnung des Lautstärkepegels und der Lautheit aus dem Geräuschspektrum. Verfahren nach E. Zwicker. Normenausschuss Akustik. Berlin, 1991
- [11] *DIN 45631/A1*: Berechnung des Lautstärkepegels und der Lautheit aus dem Geräuschspektrum. Verfahren nach E. Zwicker. Normenausschuss Akustik. Berlin, 2010
- [12] *DIN 45692*: Messtechnische Simulation der Hörempfindung Schärfe. Normenausschuss Akustik. Berlin, 2009