Signalverarbeitung für Kommunikationssysteme von Atemschutzvollmasken

Dissertation

zur Erlangung des akademischen Grades
Doktor der Ingenieurwissenschaften
(Dr.-Ing.)
der Technischen Fakultät
der Christian-Albrechts-Universität zu Kiel

vorgelegt von

Michael Brodersen

Kiel 2022

Berichterstatter:	Prof. Dr.-Ing. Gerhard Schmidt Prof. Dr. Ing. Peter Jax
Datum der mündlichen Prüfung:	23.09.2022

Kurzfassung

Im Atemschutzeinsatz ist die Kommunikation unter Feuerwehrleuten aufgrund der starken akustischen Dämpfung der Atemschutzvollmaske und der lauten Umgebungsgeräusche sehr erschwert. Um eine Kommunikation im normalen Einsatz zu gewährleisten, werden die Köpfe aneinander gehalten und die Einsatzkräfte schreien sich an. Dabei wird der Fokus vom Einsatz abgelenkt. Zur Verbesserung der Kommunikation gibt es Kommunikationssysteme, welche in die Atemschutzvollmaske integriert werden können. Diese zeichnen das Sprachsignal mit einem Mikrofon auf und geben dieses Signal an die Umgebung mit Lautsprechern, über den Team-Funk an Einsatzkollegen und über den taktischen Funk an die Einsatzleitung aus. Mit diesem Mikrofon und den Lautsprechern ist eine Verbesserung nur bedingt möglich, da laute Atem- und Umgebungsgeräusche und die entstehenden Rückkopplungen mit den lokalen Lautsprechern die Sprachverständlichkeit stark einschränken. Um eine mögliche Steigerung der Sprachverständlichkeit zu erreichen, werden in dieser Arbeit verschiedene Verfahren der Signalverarbeitung untersucht. Die bei der Kommunikationseinheit zur Verfügung stehenden Rahmenbedingungen sind ein Mikrofon und ein Lautsprecher vor der Maske, ein Lautsprecher an den Ohren, ein elektrischer Signalausgang zum taktischen Funkgerät und ein digitaler Signalprozessor. Die störenden Atemgeräusche werden in einer Sprachaktivitätsdetektion mittels Mustererkennung erkannt und gefiltert. Die Umgebungsgeräusche werden mittels einer Geräuschschätzung ermittelt und daraufhin unterdrückt. Die Verstärkung der Sprache ist aufgrund der Rückkopplungsproblematik zwischen Mikrofon und Lautsprechern limitiert, wobei eine hinreichende Verstärkung zur Steigerung der Sprachverständlichkeit durch eine Rückkopplungskompensation ermöglicht wird. Durch die hohe Korrelation des Mikrofonsignals mit dem Lautsprechersignal muss eine Signaldekorrelation vorgenommen werden. Diese Algorithmen werden auf das Mikrofonsignal angewandt, das dadurch verbesserte Signal wird auf die jeweiligen Ausgangskanäle gemischt. Auf jedem Ausgangskanal erfolgt eine Nachverarbeitung, welche einen Exciter, einen Equalizer, einen Regelverstärker und einen Hard-Limiter beinhaltet. Mit dem Exciter werden mit nichtlinearen Kennlinien Signalanteile erzeugt, welche durch die Maskendämpfung verloren gehen. Mit dem Equalizer werden das Mikrofon und die Lautsprecher entzerrt und mit dem darauffolgenden Regelverstärker wird eine Dynamikanpassung des Signals während der Laufzeit durchgeführt. Mit diesen Maßnahmen wird in der Nachverarbeitung die Sprachqualität und -verständlichkeit gesteigert. Abschließend werden mit dem Hard-Limiter die Signale zum Schutz der Sicherungen der Elektronik begrenzt. Alle in dieser Arbeit beschriebenen Algorithmen sind auf einem 16-Bit Festkomma-Signalprozessor umgesetzt und hinsichtlich der Laufzeit optimiert. Abschließend wird ein mögliches Evaluierungsszenario für Maskenkommunikationssysteme vorgestellt.

Stichwörter: Atemschutzvollmasken, Sprachsignalverarbeitung, Rückkopplungskompensation, Geräuschreduktion, Sprachaktivitätsdetektion, Sprachverständlichkeitsverbesserung

Abstract

Full-face masks are essential for fire fighters to ensure respiratory protection in smoke diving incidents. While such masks are absolutely necessary for protection purposes on the one hand, they impair the voice communication of fire fighters drastically on the other hand. For this reason mask integrated communication systems can be used to amplify the speech, therefore, to improve the communication intelligibility and quality. The communication system picks up the speech signal by a microphone in the mask, enhances it by a digital signalprocesser, and plays back the amplified signal by loudspeakers located on the outside of such masks, transmits the signal via a local wireless network to other communication systems and routes the signal to an attached tactical radio. The enhancement via microphone and loudspeakers is only possible to a limited extend, due to the disturbing breathing and ambient noise, and the resulting coupling feedback of the loudspeaker to the microphone.

To increase the speech intelligibility and solve the problems shown before, this work examines different algorithms to improve communication for masks based on digital signal processing. Since breathing noise is picked up by the microphone, it is detected and suppressed by a voice activity detection. This algorithm ensures that only speech components are played back. In addition the ambient noise is estimated and suppressed. Due to the fact that the microphone is located close to the loudspeaker, feedback is occurring and this is reduced by feedback cancelation. To enhance the functionality of the canceler a decorrelation stage is applied to the signal. After the microphone enhancement the signals are mixed to the dedicated output signals. The post processing is possible for each output signal and includes an exciter, an equalizer, a dynamic range control, and a hard limiter. The exciter regenerates lost signal components due to the attenuation through non-linear characteristics. Equalization filters are applied to improve the stability of the system on the one hand and to enhance the perceived quality of the output signals on the other hand.

All described processing steps are implemented on a 16-bit fixed point digital signal processor and optimized for efficiency. Finally possible evaluation scenarios for masks communication system are presented.

Keywords: Full-face mask communication, feedback cancellation, noise reduction, voice activity detection, speech intelligibility enhancement

Inhaltsverzeichnis

Abbildungsverzeichnis xi

Tabellenverzeichnis xiii

Abkürzungen und Notation xiv

1 Einleitung 1
- 1.1 Lösungsansatz und Zielsetzung . . . 1
- 1.2 Aufbau der Arbeit . . . 2
- 1.3 Eigene Veröffentlichungen zur vorliegenden Arbeit . . . 3

2 Eigenschaften von Atemvollschutzmasken und Kommunikationseinheiten 5
- 2.1 Eigenschaften von Atemschutzvollmasken . . . 5
- 2.2 Eigenschaften von Kommunikationseinheiten für Atemvollschutzmasken . . . 8
  - 2.2.1 Funkgerätekommunikation . . . 11
- 2.3 Verwendete Hardware . . . 12

3 Übersicht der Signalverarbeitung 15
- 3.1 Funkgeräte-Signalverbesserung . . . 16
- 3.2 Mischung und Verstärkung der Signale . . . 17

4 Mikrofon-Signalverbesserung 19
- 4.1 Vorverarbeitung . . . 20
  - 4.1.1 Hochpassfilterung . . . 20
  - 4.1.2 Analyse- und Synthese-Filterbank . . . 21
  - 4.1.3 Kurzzeit-Fouriertransformation der Filterbänke . . . 21
- 4.2 Sprachaktivitätserkennung . . . 22
  - 4.2.1 Problemstellung . . . 23
  - 4.2.2 Merkmalsextraktion . . . 25
  - 4.2.3 Lineare Diskriminanzanalyse . . . 30
  - 4.2.4 Mustererkennung . . . 31
  - 4.2.5 Performance-Vergleich . . . 41
- 4.3 Rückkopplungskompensation . . . 45
  - 4.3.1 Vergleich der Kompensationsansätze . . . 46
  - 4.3.2 Blockverarbeitungskompensation . . . 48
  - 4.3.3 Schrittweitenkontrolle . . . 50
  - 4.3.4 Signaldekorrelation . . . 53
  - 4.3.5 Verzögerung . . . 58
- 4.4 Geräuschunterdrückung . . . 60
  - 4.4.1 Geräuschschätzung . . . 60

4.4.2 Wiener-Filter 61

5 Nachverarbeitung **65**
5.1 Exciter 66
5.1.1 Höhen-Exciter 68
5.1.2 Mitten-Exciter 73
5.2 Equalizer-Filter 77
5.2.1 Verwendete Equalizer 78
5.3 Regelverstärker 79
5.3.1 Verwendete Dynamik-Kennlinien 82
5.3.2 Anwendung der Dynamikanpassung am Beispiel der VA-Kennlinie . 83

6 Evaluierung **87**
6.1 Charakteristische Merkmale der Atemvollschutzmaske für die Evaluierung . 87
6.2 Subjektiver Hörtest des VA's 88
6.2.1 Übertragungscharakteristik 88
6.2.2 Subjektiver Hörtest 88
6.3 Evaluierungssystem und -szenarien 91
6.3.1 Eigenwahrnehmung/-störung 92
6.3.2 Kommunikation mittels der VA-Lautsprecher 93
6.3.3 Kommunikation mittels des Team-Funks 93
6.3.4 Kommunikation mittels des taktischen Funks 94
6.3.5 Bewertung der Evaluierung 95

7 Zusammenfassung und Ausblick **99**
7.1 Zusammenfassung 99
7.2 Ausblick 100

Literaturverzeichnis **101**
Publikationen mit Eigenbeteiligung 101
Weitere Literatur 101

8 Anhang **109**

Abbildungsverzeichnis

2.1 Mechanische Atemvollschutzmaske (Modell *Panorama Nova*, Dräger [Pan]). 5
2.2 Atemschutzeinsatz mit Pressluftatmer und Lungenautomat 6
2.3 Amplitudengang der Atemvollschutzmaske *Panorama Nova*. 6

3.1 Signalverarbeitungsübersicht. 15
3.2 Funkgeräte-Signalverbesserung. 16

4.1 Mikrofon-Signalverbesserung. 19
4.2 Spektrogramm vom Mikrofonsignal einer Kommunikationseinheit. 24
4.3 Signalflussgraph der Sprachaktivitätserkennung. 24
4.4 Signalflussgraph der Merkmalsextraktion. 25
4.5 Mel-Filterung mit 10 Stützstellen. 26
4.6 Genäherte Mel-Filterung mit 10 Stützstellen. 27
4.7 Logarithmus dualis. 28
4.8 Näherung des Logarithmus dualis. 29
4.9 Signalflussgraph des Codebuchs einer Klasse. 34
4.10 Beispielhaftes neuronales Netz. 35
4.11 Sigmoid Übertragungsfunktion. 37
4.12 Verschiedene Rückkopplungszweige (verändert nach [Rey+11]). 38
4.13 Verwendetes neuronales Netz. 41
4.14 Spektrogramm der Sprachaktivitätsdetektion 44
4.15 Breitbandige Anregung zur Analyse der Ausschwingzeit. 45
4.16 Struktur der Rückkopplungskompensation mit einer Teilbandverarbeitung . 47
4.17 *Normalized Fast Block LMS* (angelehnt an [Hän+04]). 49
4.18 Darstellung der Schrittweite μ_{init} mit dem Parameter $\gamma = 0,7$ 51
4.19 Schrittweite während der Laufzeit ν_{run} mit $\gamma = 0,7$ und $\nu_{\text{init}} = 0,5$. 52
4.20 Glättungskonstante γ mit $\nu_{\text{run}} = 0,2$ und $\nu_{\text{init}} = 0,5$. 52
4.21 Kohärenzanalyse der Frequenzverschiebung 56
4.22 Darstellung der statistischen Werte in einem Boxplot (nach: [Fah+16]) . . 58
4.23 Testergebnisse im Zusammenhang mit der Frequenzverschiebung. 59
4.24 Kohärenzanalyse bei einer Systemverzögerung. 60
4.25 Spektrogramm der Geräuschunterdrückung 63

5.1 Signalflussgraph der Nachverarbeitung. 66
5.2 Vergleich zweier Spektrogramme (klares Sprachsignal; mit dem Maske). . . 67
5.3 Übersicht des Exciter-Signalflussgraphen (verändert nach: [Bul+16]) 68
5.4 Vollständiger Exciter-Signalflussgraph (nach: [Bul+16]) 68
5.5 Kennlinie K für die Erzeugung der geraden und ungeraden Harmonischen. 69
5.6 CMOS-Testergebnisse im Zusammenhang mit dem Höhen-Exciter. 70
5.7 Signal nach Bearbeitung mit dem Höhen-Exciter. 71

5.8 Kohärenzanalyse bei einem Höhen-Exciter. 72
5.9 Signalflussgraph des Mitten-Exciters 73
5.10 Frequenzverlauf des IIR-Butterworth-Bandpasses $H_{ex,mid}$ 73
5.11 Kennlinie der geraden Harmonischen für den Mitten-Exciter. 74
5.12 CMOS-Testergebnisse im Zusammenhang mit dem Mitten-Exciter 75
5.13 Signal nach Bearbeitung mit dem Mitten-Exciter. 76
5.14 Kohärenzanalyse bei einem Mitten-Exciter. 76
5.15 Frequenzgang des Equalizers vom taktischen Funkausgangssignal 78
5.16 Frequenzgang des Equalizers vom VA-Signal 79
5.17 Beispiel einer Dynamik-Kennlinie 80
5.18 Beispiel einer Dynamik-Kennlinie mit *hard knee*- und *soft knee*-Übergang . 81
5.19 Signalflussgraph des Regelverstärkers 81
5.20 Dynamik-Kennline der Ohrlautsprecher 83
5.21 Dynamik-Kennline der VA-Lautsprecher 83
5.22 Beispielhafte Dynamikanpassung mit der Dynamik-Kennlinie. 84

6.1 Übertragungscharakteristik der Kommunikationseinheit 89
6.2 Evaluierungsaufbau des MRT 90
6.3 *Head and torso simulators* aus unseren Evaluierungsexperimenten. 91
6.4 Überblick der Evaluierungsszenarien für Kommunikationssysteme. 92
6.5 Beispielhaftes Qualitätsdiagramm einer Kommunikationseinheit 96

8.1 IIR-Filter erster Ordnung 109
8.2 IIR-Filter 2. Ordnung in der Direkt-Form 2 109

Tabellenverzeichnis

2.1 Rechenzyklen der FFT auf dem Ko- gegenüber dem Hauptprozessor. 13

4.1 Erkennungsmatrix des neuronalen Netzes. 41
4.2 Erkennungsmatrix des Codebuches. 42
4.3 Benötigte Ressourcen der Mustererkenner. 42
4.4 Vor- und Nachteile der verschiedenen Kompensationsansätze. 46
4.5 Beispielhafte MOS-Skala bei der Bewertung der Sprachqualität. 56
4.6 Beispielhafte CMOS-Skala bei der Bewertung der Sprachqualität. 57
4.7 MOS-Skala bei der Bewertung der Sprachverständlichkeit. 57

Abkürzungen

A/D-Wandler	Analog-Digital-Wandler
AGC	*Automatic gain control*, Automatische Verstärkungssteuerung zum Ausgleich unterschiedlicher Sprecherlautstärken
CMOS	*Comparison mean opinion score*
CSA	Chemikalienschutzanzug
D/A-Wandler	Digital-Analog-Wandler
DFT	Diskrete Fourier-Transformation
DMA	*Direct memory access*, Speicherdirektzugriff
DSP	Digitaler Signalprozessor
DUT	*Device unter test*
FFT	*Fast fourier transform*, Schnelle Fourier-Transformation
FIR	*Finite impulse response*, Filter mit endlicher Impulsantwort
HATS	*Head and torso simulator*
IDFT	Inverse diskrete Fourier-Transformation
IIR	*Infinite impulse response*, Filter mit unendlicher Impulsantwort
I2S	*Inter-IC sound*
LBG	*Linde-Buzo-Gray*
LDA	Lineare Diskriminanzanalyse
LMS	*Least mean squares*
MAC	*Multiply-accumulate*, Multiplizier- und Akkumuliereinheit
MFCC	*Mel frequency cepstral coefficients*, Mel-Frequenz-Cepstrum-Koeffizienten
MOS	*Mean opinion score*
MRT	*Modified rhyme test*
MSE	*Mean-squared error*, mittlerer quadratischer Fehler
NDGC	*Noise dependent gain control*, Geräuschabhängige Verstärkungssteuerung
NLMS	*Normalized least mean squares*
OLA	*Overlap-add*
POLQA	*Perceptual objective listening quality assessment*
PTT	*Push-to-Talk*
SCBA	*Self-contained breathing apparatus*, Pressluftatmer
SD	spektrale Distanz
SNR	*Signal-to-noise ratio*, Signal-Rausch-Verhältnis
SOS	*Second order structure*
SPL	*Sound pressure level*, Schalldruckpegel
STFT	*Short-time fourier transform*, Kurzzeit-Fourier-Transformation
STI	*Speech transmission index*

TETRA	*Terrestrial trunked radio*
TI	*Texas Instruments*
VA	*Voice amplifier*, Sprachverstärker
VAD	*Voice activity detection*, Sprachaktivitätsdetektion

Notation

Konventionen und Operatoren

j	Imaginäre Einheit, $j^2 = -1$
$\Re\{z\}$	Realteil der komplexen Zahl z
$\Im\{z\}$	Imaginärteil der komplexen Zahl z
z	Komplexe Zahl, $z = \Re\{z\} + j\,\Im\{z\}$
$\lvert z\rvert$	Betrag
$\angle z$	Argument/Phasenwinkel
z^*	Zu z konjugiert komplexe Zahl, $z^* = \Re\{z\} - j\,\Im\{z\}$
$\mathbf{x} = (x_i)$	Vektor $\mathbf{x}$ mit Elementen x_i
$\mathbf{x}^{\mathrm{T}}$	Transponierter Vektor
$\lVert\mathbf{x}\rVert_p$	p-Vektornorm (Euklidische Norm für $p = 2$), $\lVert\mathbf{x}\rVert_p = \left(\sum_{i=0}^{N-1} \lvert x_i\rvert^p\right)^{1/p}$
$\mathbf{A} = (A_{i,j})$	Matrix $\mathbf{A}$ mit Elementen $A_{i,j}$
$\mathbf{A}^{\mathrm{T}}$	Transponierte Matrix
$\mathbf{A}^{-1}$	Inverse Matrix
$\mathbf{A}^{\dagger}$	Pseudoinverse Matrix
$\mathbf{A}^{\mathrm{H}}$	Hermitesche Matrix, $\mathbf{A}^{\mathrm{H}} = (\mathbf{A}^*)^{\mathrm{T}}$
$\mathbb{N}$, $\mathbb{Z}$, $\mathbb{R}$, $\mathbb{C}$,	Menge der natürlichen, ganzen, rationalen, komplexen Zahlen
$\cup_{g=1}^{N} \mathbb{M}_i$	Vereinigung von N Mengen $\mathbb{M}_i$
$\lceil\cdot\rceil$, $\lfloor\cdot\rfloor$, $\lfloor\cdot\rceil$	Aufrunden, abrunden, runden auf die nächste ganze Zahl
$[a, b]$, $[a, b)$	Geschlossenes, halboffenes Intervall
$\arg\max\{f(x)\}$	Wert x für das f sein Maximum annimmt
$\mathrm{E}\{x\}$	Erwartungswert der Zufallsvariable x
$\mathrm{Var}\{x\}$	Varianz der Zufallsvariable x
$*$	Faltungsoperator
$\mathrm{DFT}_N\{\cdot\}$	DFT der Ordnung N
$\mathrm{IDFT}_N\{\cdot\}$	IDFT der Ordnung N
$\hat{x}$	Schätzwert von x

Symbolverzeichnis

B	Kurzzeit-Spektrum des Hintergrundgeräusches
B_{A}	Bias der Ausgangsschicht des Neuronalen Netzes
B_{V}	Bias der verdeckten Schicht des Neuronalen Netzes
$\mathbf{c}$	Codebucheinträge
$d_{\mathrm{E},b}$	Quadratische euklidische Distanzen der Klassen des Codebuchs
d_{min}	Ergebnis des Codebuchs
d_{vad}	Ergebnis des Mustererkenners der VAD
e	Signal nach der Rückkopplungskompensation
E	Kurzzeit-Spektrum nach der Rückkopplungskompensation
e_{m}	Mittelwert der quadratischen euklidischen Distanz des Codebuchs
E_{s}	Kurzzeit-Spektrum nach der Geräuschunterdrückung
f	lineare Frequenz
f_{s}	Abtastrate
f_{shift}	Frequenzversatz in Hz
$\hat{F}$	Kurzzeit-Spektrum der geschätzten Rückkopplung
g_{dyn}	Verstärkungsfaktor des Regleverstärkers
g_h	Verstärkungsfaktor des Exciters
h_{ana}	Analysefenster der Filterbank
h_{syn}	Synthesefenster der Filterbank
h_{win}	Fensterfunktion
h_{hanning}	Hann-Fenster
H	Übertragungsfunktion
k	Rahmenindex
K	Menge aller Klassen der VAD
K_{g}	Kennlinie zur Erzeugung der geraden Harmonischen
K_{u}	Kennlinie zur Erzeugung der ungeraden Harmonischen
L_{s}	Lautstärke
n	Diskreter Zeitindex
N_{B}	Anzahl der Codebucheinträge
N_{DFT}	DFT-Ordnung
N_{fft}	FFT-Ordnung
N_{m}	Anzahl der Merkmale der VAD
N_{LDA}	Dimension nach der LDA
N_{Sub}	Anzahl der Teilbänder
p_{eff}	Effektiver Schalldruck
$p_{\mathrm{eff},0}$	Bezugsschalldruck
R	Rahmenversatz
R_{Lsp}	Rahmenversatz des Lautsprechersignals
r	Empfangssignal des taktischen Funkgeräts
r_{e}	Signal nach der Funkgerätesignalverbesserung
S	Kurzzeit-Spektrum des Sprachsignals
S_{xx}	Autoleistungsspektrum
S_{yx}	Kreuzleistungsspektrum

v	Ergebnis der Sprachaktivitätsdetektion
x	Mikrofonsignal
X	Kurzzeit-Spektrum des Mikrofonsignals
X_e	Kurzzeit-Spektrum nach der Signaldekorrelation
x_e	Ausgangssignal der Mikrofonsignalverbesserung
$\mathbf{X}_\mathrm{LDA}$	Merkmalsvektor der VAD nach der LDA
X_dB	Logarithmierte Merkmale der VAD
$\overline{X}_{feat}$	Mittelwert der Merkmale der VAD
$\tilde{\boldsymbol{X}}$	Merkmalsvektor der VAD
y	Ausgangssignal zum Lautsprecher
y_dyn	Ausgangssignal des Regelverstärkers
y_fil	Hochpassgefiltertes Signal im Exciter
y_ex	Signal nach dem Exciter
Y_c	Kurzzeit-Spektrum des verzögerten Lautsprechersignals
$\boldsymbol{\Sigma}$	Kovarianzmatrix
α_attack	Ansprechzeit des Regelverstärkers
α_sm	Glättungskonstante
α_release	Rücklaufzeit des Regelverstärkers
γ	Glättungskonstante des Rückkopplungskompensators
$\gamma_{X,Y}$	Kohärenz
μ	Teilbandindex
μ_opt	Optimale Schrittweite des Rückkopplungskompensators
μ_step	Schrittweite des Rückkopplungskompensators
$\boldsymbol{\Theta}_\mathrm{LDA}$	Transformationsmatrix der LDA
Δ_inc	Steigungskonstante
Δ_dec	Abfallkonstante
τ	Timbre-Faktor

Kapitel 1

Einleitung

Während Einsätzen der Feuerwehr unter schwerem Atemschutz kommen brenzliche Situationen häufig vor. Es wird in toxischen, verrauchten und heißen Umgebungen gearbeitet, wodurch der körperliche Stress sehr hoch ist. Bei den Einsätzen ist die Sprachkommunikation sehr wichtig, so dass sich beispielsweise die Einsatzkräfte untereinander abstimmen können. Dazu zählt beispielsweise auch die Kommunikation zur Einsatzleitung außerhalb des Einsatzes. Durch die Ausrüstung, maßgeblich die Atemschutzvollmaske, wird die Kommunikation stark eingeschränkt, da die Maske die Sprache des Trägers sehr stark dämpft. Diese Dämpfung in Verbindung mit den im Einsatz vorkommenden lauten Umgebungsgeräuschen führt zu einer Beeinträchtigung der Kommunikation. Für die Kommunikation unter den Einsatzkräften bedeutet dies eine Ablenkung vom Einsatz und ggf. eine Gefahr durch falsch verstandene Anweisungen. Die Kommunikation zur Einsatzleitung erfolgt mit einem taktischen Funkgerät, welches vor die Maske gehalten wird. Durch die Positionierung des Funkgerätes überlagern die störenden Umgebungsgeräusche die Sprache. Die eingehenden Funksprüche werden über den im Funkgerät integrierten Lautsprecher ausgegeben; dieses Signal wird ebenfalls durch die Umgebungsgeräusche überlagert und die Verständlichkeit sinkt in beiden Kommunikationspfaden. Durch eine fehlerhafte Kommunikation können im Einsatz schwerwiegende Fehlentscheidungen getroffen werden, so dass beispielsweise nicht alle Verletzten rechtzeitig geborgen werden können.

1.1 Lösungsansatz und Zielsetzung

Zur Verbesserung dieser Situationen können Kommunikationssysteme an die Maske montiert werden. Diese Systeme existieren als rein passive Headsets, welche die Kommunikation mit dem taktischen Funkgerät verbessern, allerdings werden immer noch störende Umgebungs- und laute Atemgeräusche, welche durch das Atmen mit dem auf dem Rücken getragenen Pressluftatmer entstehen, überlagert. Die Kommunikationssysteme an den Masken können auch aktive Systeme sein, welche über einen digitalen Signalprozessor verfügen und somit mehr Möglichkeiten der Signalverarbeitung bieten. Diese Kommunikationssysteme haben ein Mikrofon an der Maske, welches die Sprache aufnimmt, verarbeitet und verstärkt auf den Ausgangskanälen wiedergibt. Diese Ausgangskanäle sind Lautsprecher vor der Maske, ein Lautsprecher an den Ohren des Trägers, eine elektrische Schnittstelle zum taktischen Funkgerät und ein Kurzstrecken-Team-Funk.

Das Mikrofonsignal nimmt die Sprache außerhalb der Maske vor der sogenannten Sprechmembran auf. Die Sprechmembran ist dabei ein Resonator mit einer Resonanzfrequenz von ca. 800 Hz. Durch die Schwingung der Membran ergibt sich ein erhöhter Nachhall im Mikrofonsignal und durch die Dämpfung der Maske sind nur die Frequenzen des Resonators

vollständig vorhanden. Zusätzlich existieren im Mikrofonsignal die lauten Atemgeräusche des Pressluftatmers und die Umgebungsgeräusche. Zur Verbesserung des Mikrofonsignals werden die Atem- und Umgebungsgeräusche geschätzt und unterdrückt. Zur Entzerrung des Sprachsignals werden in dieser Arbeit diverse Signalverarbeitungsalgorithmen zur Verbesserung der Sprachverständlichkeit angewendet.

Die vorderen Lautsprecher sind zur Verständigung im näheren Umfeld gedacht und daher vor der Maske in unmittelbarer Nähe zum Mikrofon angeordnet. Durch diese Positionierung ergibt sich eine elektro-akustische Rückkopplung, wodurch die Verstärkung des Mikrofonsignals limitiert ist. Um diese Verstärkung zu maximieren, wird in dieser Arbeit eine Rückkopplungskompensation mit der zugehörigen Dekorrelation vorgestellt. Die Dekorrelation des Lautsprechersignals ist notwendig, da das Mikrofonsignal stark mit dem Lautsprechersignal korreliert. Zusätzlich werden die Lautsprecher entzerrt und die Dynamik mit einem Regelverstärker angepasst, so dass eine möglichst gute Sprachverständlichkeit in der näheren Umgebung vorhanden ist. Die Lautsprecher an den Ohren sind zur Wiedergabe der eingehenden Funksprüche des taktischen Funkgeräts und des Team-Funks. Die Ohrlautsprecher werden im Rahmen dieser Arbeit wie die vorderen Lautsprecher entzerrt und in der Dynamik angepasst.

Durch die elektrische Schnittstelle zum Funkgerät ist die Kommunikationseinheit ein Headset zu diesem. Damit verbunden ergeben sich einige Herausforderungen, da an dieser Schnittstelle sehr viele Funkgeräte von verschiedenen Herstellern angeschlossen werden können. Die verschiedenen Funkgeräte haben verschiedene Audio-Codecs, wodurch die Signalverarbeitung der Kommunikationseinheit auf diese abgestimmt sein muss. Der elektrische Stecker ist allerdings für alle Funkgeräte gleich, wodurch eine Unterscheidung in der Kommunikationseinheit nicht möglich ist. Daher ist es Teil dieser Arbeit die Signalverarbeitung dementsprechend zu untersuchen und entwerfen, ob ein allgemeingültiger Ansatz für die verschiedenen Audio-Codecs besteht. Untersucht werden dafür Filter und Dynamikanpassungsalgorithmen.

Damit die Wirkungsweise der Algorithmen dargestellt werden kann, werden im Rahmen dieser Arbeit zu den verschiedenen Algorithmen Hörversuche und Untersuchungen durchgeführt, so dass beispielsweise eine Steigerung der Sprachverständlichkeit belegt werden kann. Ein Problem in Bezug auf die Kommunikation ist die Bewertung der Systeme: ab wann ist ein System gut und bringt den Einsatzkräften eine wirkliche Verbesserung? Um diese Bewertung in der Zukunft möglich zu machen, werden im Rahmen dieser Arbeit ein Evaluierungssystem und verschiedene Evaluierungsszenarien vorgestellt.

1.2 Aufbau der Arbeit

Die Dissertation strukturiert sich wie folgt:

Das Kapitel 2 stellt zunächst Eigenschaften der Atemschutzvollmasken vor und geht auf die Vorteile und Probleme ein. Daraufhin werden in diesem Kapitel die Eigenschaften der Kommunikationssysteme vorgestellt. Dabei werden deren Komponenten, Schnittstellen und Herausforderungen dargestellt. Bei den Komponenten wird die verwendete Hardware hinsichtlich des verfügbaren Energiebudgets und des Prozessors erläutert. Anschließend gibt das Kapitel 3 eine komplette Übersicht über die Signalverarbeitung. Im Kapitel 4 werden die Algorithmen zur Verbesserung des Mikrofonsignals erläutert, welche eine Analyse- und Synthesefilterbank, eine Sprachaktivitätsdetektion zur Filterung der Atemgeräusche,

eine Geräuschunterdrückung der Hintergrundgeräusche und eine Rückkopplungskompensation mit der zugehörigen Dekorrelation des Lautsprechersignals beinhalten. Das Kapitel 3.2 beschreibt die Mischung und Verstärkung der Signale, wobei die Eingangssignale auf die Ausgangssignale gemischt und verstärkt werden. Zusätzlich werden die notwendigen zu beachtenden Regeln bei der Priorisierung der Signale vorgestellt. Das Kapitel 5 schildert die Nachverarbeitung für die Ausgangskanäle, welche einen Exciter zur Reproduktion von Sprachanteilen, einen Equalizer zur Entzerrung der Signale, einen Regelverstärker zur Anpassung der Dynamik und einen Hard-Limiter zur Begrenzung der Signale beinhaltet. Das Kapitel 6 setzt sich mit möglichen Evaluierungssystemen und -szenarien auseinander, bevor das Kapitel 7 mit einer Zusammenfassung und einem Ausblick die Arbeit abschließt.

1.3 Eigene Veröffentlichungen zur vorliegenden Arbeit

Dieses Kapitel stellt die Verbindung von eigenen Veröffentlichungen mit Bezug auf die Kommunikationssysteme von Atemschutzvollmasken zur vorliegenden Arbeit dar. Die Erstautorenschaft ist hierbei mit * gekennzeichnet. Die Sprachaktivitätsdetektion der Kommunikationseinheit wird in der ersten Veröffentlichung [Bro+15]* beschrieben. In dieser wird die Merkmalsextraktion und die Mustererkennung zur Unterscheidung von Sprache, Atemgeräuschen und Umgebungsgeräuschen untersucht und die Implementierung beschrieben. Die Sprachaktivitätsdetektion ist ein zentrales Element der Signalverarbeitung, da mit dieser sehr viele Kanäle in der Kommunikationseinheit gesteuert werden. Anschließend wurde versucht, die Sprachverständlichkeit durch eine geeignete Signalverarbeitung zu steigern [Grö+17]. Hierbei werden mittels nichtlinearer Kennlinien Harmonische der Grundfrequenz auf den Ausgangssignalen erzeugt. Die Harmonischen sind durch die Dämpfung der Maske verloren gegangen. Zur Darstellung der Steigerung der Sprachverständlichkeit sind subjektive Hörtests durchgeführt worden. Aufbauend auf diesen Veröffentlichungen ist ein Journal Beitrag [Bro+19]* mit einer Übersicht der Signalverarbeitung der Kommunikationseinheit publiziert worden. Die Übersicht beinhaltet eine Beschreibung der akustischen Eigenschaften der Masken und den akustischen Aufbau der Kommunikationseinheiten. Daraufhin wird die Signalverarbeitung beschrieben, welche die Filterbänke, die Sprachaktivitätsdetektion, die Rückkopplungskompensation inklusive Dekorrelation, die Störgeräuschunterdrückung, die Entzerrung und die Pegelanpassung der verschiedenen Signalpfade umgesetzt. Somit gibt diese Veröffentlichung bereits eine sehr gute Übersicht über die gesamte Signalverarbeitung und deren Herausforderungen. Zur weiteren Überprüfung der Verbesserung der Sprachverständlichkeit des Gesamtsystems ist ein subjektiver Hörversuch mit der zugehörigen Auswertung enthalten. Damit in Zukunft verschiedene Kommunikationssysteme gut evaluiert werden können, ist in der Veröffentlichung [Bro+16]* eine Übersicht über mögliche Evaluierungsszenarien mit dessen Bewertung enthalten. Darüber hinaus wurde in der Veröffentlichung [Mar+17] die Auswirkung des kombinierten Effektes von physischen und kognitiven Stress in Verbindung mit der Atemschutzvollmaske auf die Sprache untersucht.

Kapitel 2

Eigenschaften von Atemvollschutzmasken und Kommunikationseinheiten

2.1 Eigenschaften von Atemschutzvollmasken

Atemschutzvollmasken schützen das Gesicht des Trägers sowie dessen Atemwege vor toxischen Gasen und Rauch (siehe [Vol+13]). Die Atemschutzvollmasken werden im Folgenden zur Vereinfachung Maske genannt. Die Maske wird durch eine Dichtlinie um das Gesicht abgedichtet, Nase und Mund werden von der Innenmaske bedeckt, welche die Ausatemluft so lenkt, dass das Visier nicht beschlägt (siehe Abb. 2.1). Der Raum vor Mund und Nase wird mit frischer Luft aus einem Lungenautomaten, der an einen Pressluftatmer angeschlossen ist, versorgt. Die Luft kommt aus der auf dem Rücken getragenen Flasche des Pressluftatmers (siehe Abb. 2.2). Im Lungenautomaten wird ein Ventil geöffnet, sobald der Träger einatmet; beim Ausatmen schließt es wieder und die Ausatemluft strömt durch ein andes Ventil unterhalb des Lungenautomaten aus der Maske heraus.

Abbildung 2.1: Mechanische Atemvollschutzmaske (Modell *Panorama Nova*, Dräger [Pan]).

Durch die Abdichtung der Maske am Gesicht wird die Sprache stark gedämpft. Um diese Dämpfung zu minimieren, ist vor dem Mund eine Sprechmembran angeordnet, welche wie ein Resonator wirkt. Die Resonanzfrequenz bei der Panorama Nova ist bei ca.

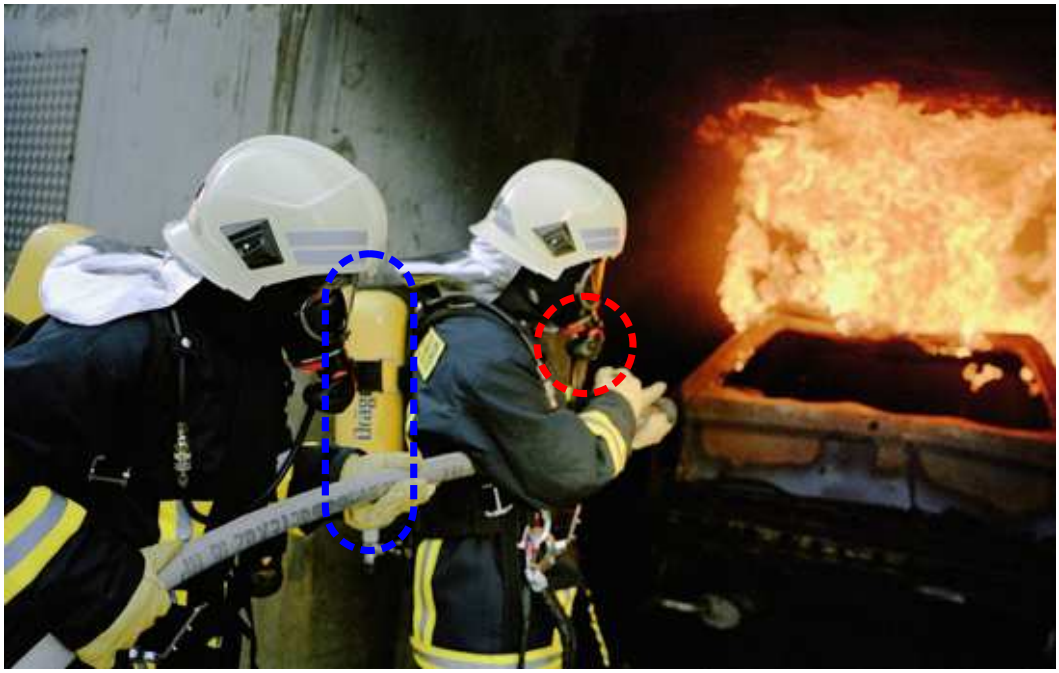


Abbildung 2.2: Atemschutzeinsatz (Lungenautomat in Rot, Flasche des Pressluftatmers in Blau).

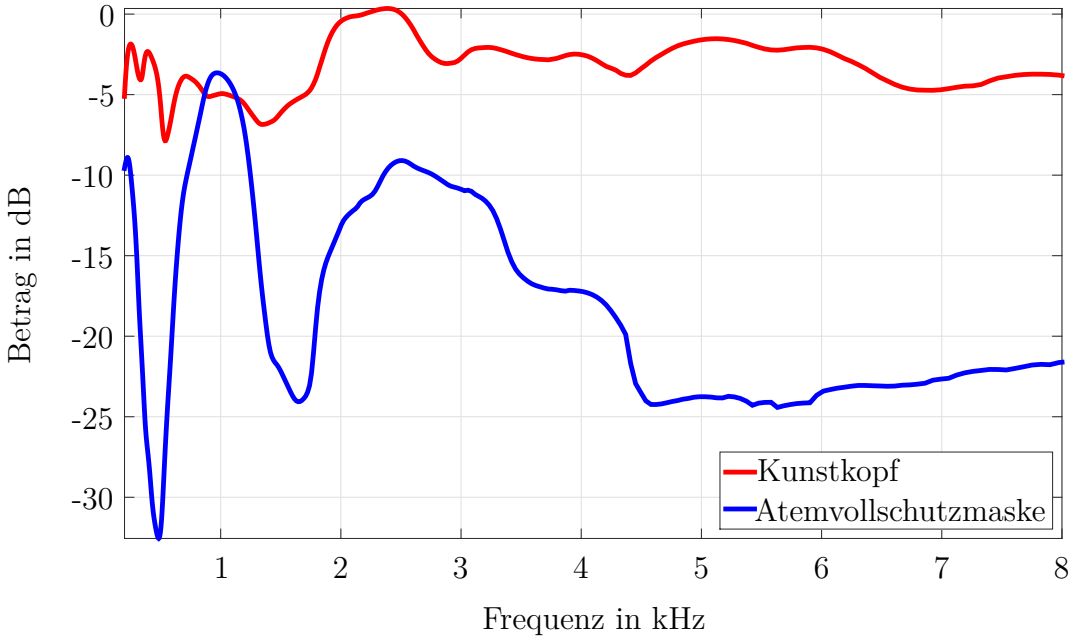


Abbildung 2.3: Frequenzgang der Atemvollschutzmaske *Panorama Nova* (blau) und der Kunstkopf ohne Atemvollschutzmaske (rot).

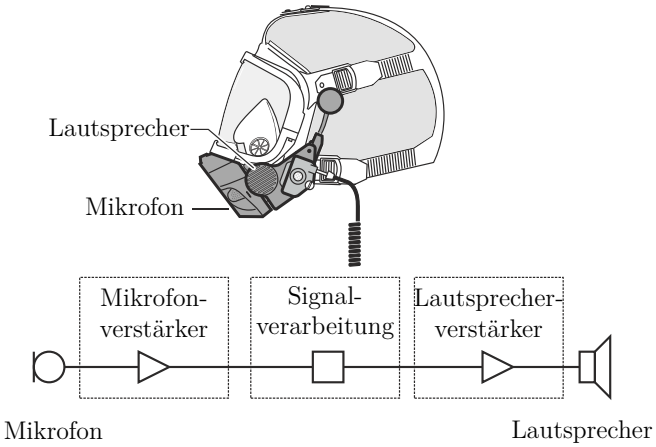
1000 Hz. Die Resonanz der Sprechmembran wird durch ihren Durchmesser verändert; je größer der Durchmesser der Membran ist, desto niedriger ist die Resonanzfrequenz und je kleiner der Durchmesser wird, desto höher wird die Resonanzfrequenz. Durch diesen Resonator hat das System für die Sprachsignale eine lange Nachhallzeit. Außerdem ist

der Vokaltrakt des Trägers durch das kleine abgeschlossene Volumen vor dem Mund verstimmt [Hub04, S. 7 f]. Die Sprechmembran besteht typischerweise aus einer dünnen Folie aus Edelstahl oder Polyimid, welche in einem Ring gespannt wird und damit eine akustische Übertragung ermöglicht. Somit können Sprachsignale die hermetische Abdichtung der Maske durchdringen, allerdings wird das Sprachsignal durch die Membran und die Maske verzerrt. Außerdem sind die Schalldruckpegel (*Sound Pressure Level*, SPL) an der Sprechmembran sehr hoch, da der Mund sehr dicht an der Membran liegt, wodurch die Membran in einem nicht-linearen Bereich betrieben werden kann. Die Übertragungsfunktion einer solchen Maske ist in der Abb. 2.3 zusehen, wobei die Frequenzen bei 500 Hz und über 1 kHz stark gedämpft und verzerrt sind, wodurch die Sprachverständlichkeit gegenüber einem Sprecher ohne Maske schlechter ist. Die Frequenzen bei der Messung der Maske unter 250 Hz sind Messfehler, hier hat die Maske normalerweise ebenfalls eine deutlich stärkere Dämpfung. Der Messfehler resultiert aus dem zur Verfügung stehenden reflexionsarmen Raum, welcher diese Frequenzen nicht dämpft.
Die mechanischen Anforderungen an die Sprechmembran und an die Maske müssen den Atemschutz bei hoher chemischer, starker Rauch-, Hitze- und Wasserbelastung garantieren, welche aus den Aufgabenfeldern Brand- und Gefahrstoffeinsatz im Bereich der Feuerwehr und der Industrie resultieren. Die Belastung für die mechanischen Bestandteile ist ebenfalls bei der Reinigung nicht zu unterschätzen, da die Masken oft in Waschmaschinen gewaschen und dabei in der Trommel stark belastet werden.

2.2 Eigenschaften von Kommunikationseinheiten für Atemvollschutzmasken

Der Einsatz unter schwerem Atemschutz findet typischerweise in einer geräuschbehafteten Umgebung mit schlechter Sicht statt, wodurch die Verständlichkeit der Sprache mit der Atemvollschutzmaske begrenzt ist. Um die Kommunikation beim Atemschutzeinsatz weiter zu verbessern, werden Kommunikationseinheiten (siehe Abb. 2.4a) verwendet, welche die Sprache mit einem Mikrofon vor der Sprechmembran außerhalb der Maske aufzeichnen, verarbeiten und verstärkt auf die Lautsprecher, ein taktisches Funkgerät oder den Team-Funk ausgeben. Dieser Signalfluss ist in Abb. 2.4b dargestellt. Die eingehenden Funksprüche werden direkt über den Ohrlautsprecher ausgegeben, damit dieser eine gute Verständlichkeit hat. Das Funkausgangssignal des taktischen Funkgeräts wird dabei nur

(a) Atemvollschutzmaske (Modell FPS 7000, Dräger) mit einer Kommunikationseinheit (Modell FPS-COM 7000, Dräger [Com]) auf einem Kunstkopf.



(b) Struktureller Überblick der Kommunikationseinheit [Vol+13].

Abbildung 2.4: Maske mit Kommunikationseinheit und strukturellem Überblick.

ausgegeben, wenn der *Push-to-Talk* (PTT) Knopf gedrückt wird. Das Funkgerät dient zur Kommunikation zwischen dem Einsatztrupp und beispielsweise der Leitzentrale oder dem Einsatzleiter. Die eingehenden Funksprüche werden auf die Ohrlautsprecher gemischt und somit kann die Kommunikation mit einem Funkgerät durch die Kommunikationseinheit deutlich vereinfacht werden. Die Kommunikationseinheit wirkt für das Funkgerät wie ein Headset mit einer Signalaufbesserung durch Signalverarbeitung. Ohne eine Kommunikationseinheit wird bei der Kommunikation mit dem Funkgerät zum Senden eines Funkspruches das Funkgerät vor die Maske gehalten, wodurch die Verständlichkeit aufgrund der Umgebungsgeräusche sehr schlecht sein kann. Dabei gibt es Funkgeräte, welche eine Signalverarbeitung besitzen und Störgeräusch-Reduktionsverfahren durchführen, aber die Vielzahl der Funkgeräte ist günstig und mit wenig Signalverarbeitung ausgestattet. Bei beiden Arten von Funkgeräten ist die Kommunikation mit der Kommunikationseinheit deutlich besser, da durch die Mikrofonposition das SNR immer besser als beim Funkgerät ist, welches im Einsatz direkt vor die Maske gehalten wird. Bei einem eingehenden Funkspruch wird das Funkgerät sehr laut eingestellt, um möglichst viel zu verstehen. Die Sprachverständlichkeit wird durch die oft sehr laute Umgebung erschwert. Bei der Kommunikationseinheit wird der eingehende Funkspruch auf den Ohrhörern wiedergegeben, wodurch die Sprachverständlichkeit durch das deutlich bessere SNR gesteigert wird.
Der Team-Funk ist ein proprietärer Funk der Kommunikationseinheit. Dieser dient zur Kommunikation im Einsatztrupp, wobei die Kommunikationseinheiten miteinander über Funk kommunizieren. Dabei wird das Mikrofonsignal zu den Kommunikationseinheiten im Einsatztrupp gesendet und bei den Empfangseinheiten auf dem Ohrhörer wiedergegeben. Somit ist die Kommunikation im Einsatztrupp auch bei einer sehr lauten Umgebung gesichert. Diese Funkkommunikation kann ebenfalls im Einsatz mit dem Chemikalienschutzanzug (CSA) [Cps] sehr hilfreich sein. Im Chemikalienschutzanzug ist jede Einsatzkraft für sich in einem abgeschlossenen Anzug, wodurch die Kommunikation mit dem Sprachverstärker (*Voice Amplifier*, VA) nicht mehr möglich ist. Bei einem CSA-Einsatz müssen sich die Personen sehr oft absprechen, um beispielsweise Chemikalien-Leckagen gemeinsam abzudichten. Daher erleichtert der Team-Funk den Einsatz sehr. Zur Aktivierung des Team-Funks wird die Sprachaktivitätserkennung (*Voice Activity Detection, VAD*) genutzt, wodurch beim Arbeiten dauerhaft beide Hände zur Verfügung stehen. Mit den genannten Kommunikationsmöglichkeiten kann der Atemschutzgeräteträger mit der Kommunikationseinheit also in der näheren Umgebung mit den Personen über den VA, mit den anderen Atemschutzgeräteträgern über den Team-Funk und mit dem angeschlossenen Funkgerät mit dem Einsatzleiter oder mit der Leitzentrale kommunizieren, wodurch sich die Kommunikation im Vergleich zu der rein mechanischen Maske deutlich vereinfacht hat.
Für eine gute Kommunikation ist es notwendig, nur das Nutzsignal und nicht das Störsignal zu übertragen, damit der Inhalt verständlicher wird. Bei der Masken-Kommunikationseinheit sind besonders die Atemgeräusche störend. Die Atemgeräusche werden durch Strömungsgeräusche im Lungenautomaten und der Maske erzeugt und sind durch die räumliche Nähe des Mikrofons, in dessen Signal mit sehr hohem Pegel vorhanden. Um diese störenden Signale nicht über die Lautsprecher oder über ein taktisches Funkgerät auszugeben, ist eine Sprachaktivitätserkennung und eine anschließende Filterung notwendig, welche zwischen den Nutz- und Störsignalen unterscheidet. Wenn die Atemgeräusche nicht gefiltert werden, ist die Verständlichkeit erschwert, beispielsweise beim taktischen

Funk sendet der Atemschutzträger der Leitzentrale einen Funkspruch, wobei der Hörer in der Leitzentrale die Lautstärke des Funkgerätes für die Sprache einstellt. Wenn die Atemgeräusche nicht gefiltert sind, wobei das Einatmen durch den Pressluftatmer einen besonders hohen Schalldruckpegel hat, muss der Zuhörer die Lautstärke immer wieder der Sprache und den lauten Störgeräuschen anpassen. Die Verständlichkeit beim Hörer sinkt somit. Im Atemschutzeinsatz kann dies ebenfalls sehr störend sein: Wenn beispielsweise mehrere Atemschutzträger mit Kommunikationseinheit in räumlicher Nähe stehen und die Atmung nicht synchron ist, überlagern die Störgeräusche permanent die Sprachsignale. Somit sinkt ebenfalls die Verständlichkeit.
Ein weiterer wichtiger Aspekt ist die Verstärkung des Mikrofonsignals auf den Lautsprechern, welches durch die Rückkopplung limitiert wird. Dies resultiert aus dem sehr kurzen Kopplungsweg vom Lautsprecher zum Mikrofon, welcher nur ca. 4 cm beträgt; somit bildet sich sehr schnell eine Rückkopplung aus. Diese Rückkopplung kann mittels Signalverarbeitung auf dem Mikrofonsignal verringert werden, damit eine höhere Verstärkung erzielt werden kann und die Verständlichkeit in lauten Umgebungen steigt.
Für die Kommunikationseinheit gelten ähnliche Anforderungen an die mechanischen Komponenten wie bei der Maske. Die Kommunikationseinheit muss Hitze-, Rauch-, chemischer und Wasserbelastung standhalten. Die Hitze- und Wasserbelastung sind speziell für die Lautsprecher, das Mikrofon und die Elektronik eine besondere Herausforderung. Das Gehäuse muss wasserdicht sein, damit die Elektronik nicht beschädigt wird. Die Lautsprecher und das Mikrofon müssen einer Handwäsche standhalten und dürfen danach keine Funktionseinschränkungen aufzeigen, wobei sehr aggressive Waschmittel verwendet werden. Die Lautsprecher dürfen dementsprechend keine empfindliche Membran besitzen und es dürfen keine wasserlöslichen Kleber beim Zusammenbau des Lautsprechers verwendet werden. Das Mikrofon wird durch eine wasserfeste Membran geschützt, wobei die Membran einen linearen Frequenzgang hat. Somit müssen alle Komponenten, Akustik, Elektronik und Mechanik, sorgfältig ausgewählt werden, um den Anforderungen gerecht zu werden.
Eine weitere Anforderung ist die Batterielaufzeit der Kommunikationseinheit. Diese wird mit zwei AA-Zellen versorgt. Mit dessen Energiekapazität muss das Gerät eine möglichst lange Laufzeit haben. Die AA-Zellen sind aufgrund des Explosionsschutzes der Kommunikationseinheit ausgewählt worden. Um den Energieverbrauch möglichst gering zu halten, werden die Lautsprecher durch einen Klasse-D-Verstärker versorgt, welcher im Vergleich zu einem Klasse A, B oder A/B Verstärker effizienter ist. Zusätzlich wird ein digitaler Signalprozessor (DSP) verwendet, welcher eine sehr geringe Leistungsaufnahme hat und eine 16-Bit-Festkomma-Recheneinheit besitzt. Die erforderliche Signalverarbeitung ist auf dem DSP implementiert; dabei sind die Algorithmen möglichst effizient gestaltet, damit die Rechenlast reduziert wird und so der Energieverbrauch niedrig ist.

2.2.1 Funkgerätekommunikation

Ein sehr wichtiger Teil der Kommunikation bei Feuerwehreinsätzen ist die Funkgerätekommunikation, wobei die Funkgeräte sich sehr stark bei den Kodierungs- und Dekodierungsverfahren sowie der Signalverarbeitung unterscheiden. Aktuell sind der analoge und digitale Funk ungefähr gleich stark verbreitet, wobei der analoge Funk in Deutschland in den nächsten Jahren weiter abnehmen wird, da innerhalb des digitalen Funks ein Kommunikationsnetz aufgebaut wurde. Beim analogen Funk ist der Pegel des Hintergrundrauschens des Kanals in der Regel etwas höher, allerdings ist die Kommunikation in Umgebungen ohne Störgeräusche ansonsten gut. In Umgebungen mit lauten Störgeräuschen kann die Kommunikationsqualität variieren, da die analogen Funkgeräte keine Signalverarbeitung besitzen. Im Vergleich ist beim digitalen Funk die Sprachverständlichkeit abhängig von dem jeweiligen Codec, da bei diesem die Bitraten sehr gering sein können.
Ein weiterer großer Unterschied ist die Zeit vom Betätigen der PTT-Taste bis zum Erhalten des Funkkanals. Beim analogen Funk ist die Zeit vom Betätigen des PTT bis zum Erhalten des Funkkanals sehr kurz und fast nicht wahrnehmbar. Beim digitalen Funk sieht dies entscheidend anders aus, dabei kann vom Betätigen des PTT bis zum Erhalten des Funkkanals bis zu 1 s vergehen und es können zwischen dem Loslassen des PTT und dem Freigeben des Funkkanals ebenso bis zu 1 s vergehen. Dies bedeutet für eine Kommunikation, dass zwischen dem Senden der Frage und dem Erhalten der Antwort von einem anderem Funkgerät bis zu 2 s Pause entstehen können, welches die Kommunikation sehr erschwert. Außerdem basieren die Kodierungs- und Dekodierungsverfahren beim digitalen Funk auf einem Sprachmodell, welches bei einer Kommunikation ohne Atemvollschutzmaske gut funktioniert; allerdings ist die Sprachqualität oft schlechter als bei einem analogem Funkgerät.
Bei den Kodierungs- und Dekodierungsverfahren muss zwischen dem beispielsweise in Europa geltenden *terrestrial trunked radio-CODEC* (*TETRA-CODEC*) (siehe [Tet]) und dem in Nordamerika und Australien geltenden Project 25-Codec von APCO International (*APCO P25-CODEC*) unterschieden werden. In anderen Ländern und herstellerspezifisch gibt es noch andere Kodierungsverfahren. Bei der Kommunikation von Funkgerät zu Funkgerät ist die Verständlichkeit und Qualität bei dem APCO P25-CODEC etwas besser. Dies belegt eine Messung des *Perceptual evaluation of speech quality* (PESQ) (siehe [IT01]) verschiedener Funkgeräte. Bei dieser Messung wurde das Funkgerät vor einem Kunstkopf platziert und am Empfangsfunkgerät die Messung durchgeführt. Hierbei wurde bei einem guten Funkgerät mit dem *APCO P25-CODEC* ein PESQ-Wert von 3,41 gemessen und bei einem guten Funkgerät mit dem *TETRA-CODEC* ein PESQ-Wert von 2,95 im gleichen Messszenario. Dieser Unterschied spiegelt sich auch in Kombination der Funkgeräte mit der Kommunikationseinheit der Maske wieder, wobei die PESQ-Werte jeweils ca. 0,5 niedriger sind. Als Referenz für PESQ-Messung kann eine ISDN Kommunikation verglichen werden, welche bei einer hervorragende Verbindung mit einem Sprachcodec gemäß G.711 einen PESQ-Wert von 4,4 erreicht. Zusätzlich hängt die Sprachverständlichkeit ebenfalls sehr stark von der Signalverarbeitung in dem jeweiligen Funkgerät ab. Wenn die Sprache durch die Sprechmembran in der Maske verfälscht wurde, ist die Qualität in dem TETRA-CODEC deutlich schlechter als bei dem APCO P25-CODEC. Dies resultiert aus den verschiedenen Sprachmodellen in den beiden Verfahren; daher kann die Verfälschung der Sprache durch die Maske sehr entscheidend sein und eine Entzerrung kann sehr wichtig

werden, um die Verständlichkeit zu erhöhen. Diese Vielfalt der Funkgeräte mit deren Kodierungsverfahren erschweren die Signalverarbeitung in der Kommunikationseinheit der Maske, da die Kodierungsverfahren auf die Signalveränderungen unterschiedlich reagieren, aber alle Funkgeräte mit dem gleichen Stecker angeschlossen werden können. Somit muss die Signalverarbeitung der Kommunikationseinheit für alle Kodierungsverfahren die bestmögliche Sprachverständlichkeit liefern.

2.3 Verwendete Hardware

Bei den Kommunikationseinheiten ist die Energieversorgung durch zwei AA-Batterien sichergestellt; die Batterielaufzeit des Gerätes soll möglichst lang sein, weshalb ein *Ultra Low Power fixed-point* DSP von Texas Instruments (TI) der Serie C55x verwendet wird. Bei diesem Prozessor ist der Stromverbrauch besonders gering. Dieser Festkomma-DSP hat eine Genauigkeit von 16Bit, wobei der Hauptprozessor zwei Multiplizier- und Akkumuliereinheiten (*multiply-accumulate*, MAC) hat. Diese MAC-Einheit kann beispielsweise zwei 16-Bit-Zahlen miteinander multiplizieren und eine 32-Bit-Addition in einem Zyklus durchführen [Pro13]. Diese Eigenschaften des Prozessors sind sehr interessant in Bezug auf die Minimierung der Rechenzyklen der Algorithmen, da die maximale Taktrate des Prozessors mit 120 MHz für das Funkprotokoll, die Systemsteuerung und die Signalverarbeitung ausreichen muss. Um die Minimierung der Zyklen bestmöglich auszunutzen, ist die Nutzung der *Compiler Intrinsics* empfehlenswert. Bei diesen kann beispielsweise eine Multiplikation von zwei 16-Bit-Zahlen, das Rechtsverschieben von 15 Bit und eine Begrenzung für das Zahlenformat in einem Rechenzyklus durchgeführt werden. Verschiedene *Compiler Intrinsics* gibt es für Addition, Subtraktion und Multiplikationsarten [C55c].

Bei der Verarbeitung in der Audiosignalverarbeitung werden viele Algorithmen im Spektralbereich durchgeführt. Hierzu ist eine Analyse- und Synthese-Filterbank notwendig. Die darin enthaltene schnelle Fourier-Transformation (*fast Fourier transform*, FFT) ist besonders rechenintensiv. Um die Rechenlast des Hauptprozessors zu entlasten, hat der verwendete DSP einen Koprozessor für die FFT [Pro13]. Dieser unterstützt FFTs von einer Ordnung 8 bis zu 1024 in der Abstufung von 2er Potenzen. Das Ergebnis wird immer in 16-Bit-Vektoren abgespeichert [C55a]. Die genutzten Zyklen der FFT auf dem Koprozessor sind deutlich effektiver als die Berechnung der FFT auf dem Hauptprozessor. Die verwendeten Zyklen auf dem Koprozessor gegenüber dem Hauptprozessor sind in der Tabelle 2.1 dargestellt.

FFT-Ordnung	FFT auf dem Koprozessor	FFT auf dem Hauptprozessor
8	130	291
16	170	241
32	321	748
64	436	1405
128	912	2798
256	1668	5947
512	3740	12736
1024	7315	27717

Tabelle 2.1: Verwendete Rechenzyklen der FFT auf dem Koprozessor gegenüber dem Hauptprozessor bei einer Taktrate von 100 MHz (angelehnt an [C55a, S. 20]).

Bei der größten FFT-Ordnung ist die Berechnung auf dem Koprozessor 3,8 mal so schnell und benötigt dabei nur 1/6 der Energie (siehe [C55a, S. 20]). Somit ist es sehr wichtig einen DSP mit einem Koprozessor für die FFT zu nutzen, da die FFT bei jedem Rahmen berechnet wird. Nähere Details zur Berechnung der FFT werden im Kapitel 4.1.2

beschrieben.

Das Mikrofonsignal wird vom verwendeten 4 mm-Kondensatormikrofon zum Audiocodec geleitet und im Audiocodec ist ebenfalls der Vorverstärker, der Analog-Digital-Wandler und Digital-Analog-Wandler enthalten. Hinzu können noch Filter erster Ordnung angewandt werden. Das digitalisierte Signal wird über einen Inter-IC Sound (I2S) Bus zum DSP übertragen. Auf dem DSP wird dieses Signal durch ein Speicherdirektzugriff (*Direct Memory Access*, DMA) in 64 *Samples*-Blöcken geschrieben. Der verwendete Audiocodec ist ein 4-Kanal-*Low-Power*-Codec von TI mit der Bezeichnung TLV320AIC34 [TFC+16]. Dieser Audiocodec hat zusätzlich noch verschiedene Filter und Mischerstufen integriert. Alle Audiokanäle außer das Funkgerätesignal sind am Audiocodec symmetrisch angebunden, so dass die kabelgebundenen Störungen reduziert werden [Wei96, S. 35 ff]. Das Funkgerätesignal ist unsymmetrisch, da die Funkgeräte ebenfalls nur über diese Anbindung verfügen. Für die Ohrlautsprecher werden die internen Class-D-Verstärker des Audio-Codecs genutzt und der Sprachverstärker wird über den Class-D-Verstärker von TI mit der Bezeichnung TPA2006D1 [Amp15] betrieben. Dieser Verstärker hat eine Leistung von 1,45 W. Diese Leistung klingt in Bezug auf den Sprachverstärker sehr gering, allerdings kann diese Leistung durch den Explosionsschutz nicht erhöht werden.

Kapitel 3

Übersicht der Signalverarbeitung

Eine Zusammenfassung der Signalverarbeitung wurde 2019 in dem EURASIP Journal *Audio, Speech, and Music Processing* veröffentlicht [Bro+19]. Die Signalverarbeitung der Kommunikationseinheit gliedert sich in drei Eingangs- und vier Ausgangspfade auf. Die Eingangskanäle sind das Mikrofonsignal, das ggf. angeschlossene Funkgerät und der Team-Funk. Die Ausgänge sind die vorderen Lautsprecher, die Ohrlautsprecher, das ausgehende Funkgerätesignal und der ausgehende Team-Funk, welches in der Abb. 3.1 zu sehen ist. Das Mikrofonsignal wird dabei in der Mikrofon-Signalverbesserung verarbeitet, wobei beispielsweise die Atemgeräusche unterdrückt werden. Als Ausgangssignal wird die Information der Sprachaktivität und das verbesserte Mikrofonsignal herausgegeben. Die Funkgeräte-Signalverbesserung verbessert das Funksignal des taktischen Funkgeräts und gibt dieses Signal aus. Der Team-Funk wird nicht weiter verarbeitet und direkt zur Mischung und Verstärkung durchgereicht. In der Mischung und Verstärkung können alle Eingänge verstärkt und gezielt auf die Ausgänge gemischt werden. Es wird beispielsweise

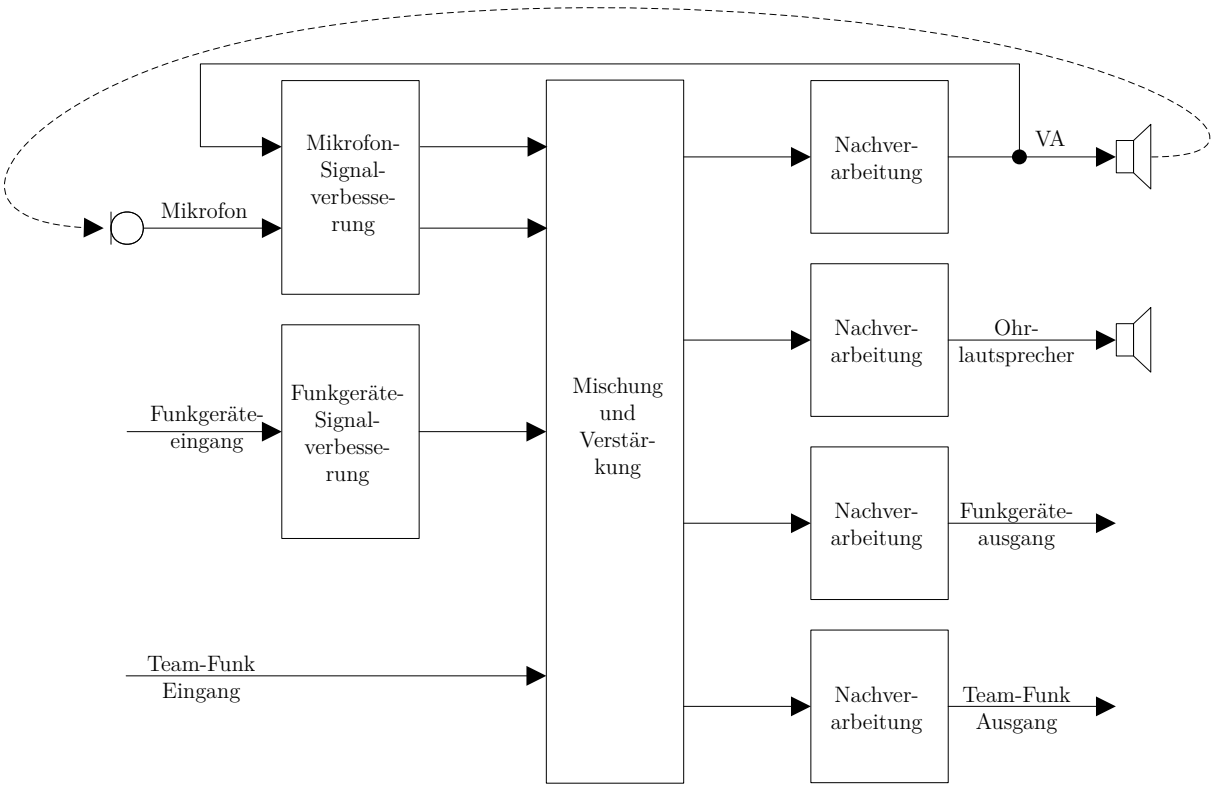


Abbildung 3.1: Signalverarbeitungsübersicht.

das Mikrofonsignal auf den VA, den Funkausgang und den Team-Funk gemischt. Dabei koppelt der VA in das Mikrofon zurück und es bildet sich eine Rückkopplungsschleife, welche in der Mikrofon-Signalverbesserung durch Algorithmen reduziert wird, um eine möglichst hohe Verstärkung zu erzielen.
Daraufhin werden das Signal des Team-Funks, das verbesserte Mikrofon- und Funksignal gemischt und verstärkt auf die jeweiligen Ausgänge verteilt. In jedem Ausgangssignal wird eine Nachverarbeitung vollzogen, in welcher ein Equalizer zur Entzerrung des Frequenzganges, ein Algorithmus zur Anpassung der Dynamik und ein Limiter zur Begrenzung der Leistung enthalten sind. Diese Nachverarbeitung wird in Kapitel 5 beschrieben. Die Nachverarbeitung findet im Zeitbereich statt, da dadurch nur in der Mikrofon- und Funkgeräte-Signalverbesserung eine Analyse- und Synthese-Filterbank gerechnet werden muss und nicht in jedem Ausgangspfad eine Synthese-Filterbank notwendig ist; somit ist der Rechenaufwand deutlich reduziert.

3.1 Funkgeräte-Signalverbesserung

Die Kommunikation mit dem Funkgerät ist essentiell bei einem Einsatz. Allerdings haben die Signale oft einen hohen Rauschpegel, wodurch ein Störpegel permanent vorhanden ist. Um eine gute Verständlichkeit zu gewährleisten, wird das Funkgeräte-Signal verbessert (siehe Abb. 3.2), wofür das Funkgeräte-Signal $r(n)$ mit dem diskreten Zeitindex n in den Frequenzbereich transformiert wird und zuvor durch die Vorverarbeitung bearbeitet und mit einem Hann-Fenster gefenstert wird; zur Vorverarbeitung dient wiederum ein Hochpassfilter wie in der Mikrofon-Signalverbesserung. Zur Reduzierung des Störpegels wird das Kurzzeit-Leistungsdichtespektrum des Geräuschs geschätzt und mit einem Wiener-Filter unterdrückt. Daraufhin wird das Signal verstärkt und ggf. entzerrt, um dessen Frequenzgang anzupassen. Diese Anpassung kann notwendig sein, falls das Funkgeräte-Signal beispielsweise durch eine Kodierung und Dekodierung verfälscht wurde. Dieses geschieht beispielsweise bei dem in Europa geltenden Tetra-Codec. Nach der Entzerrung wird das verbesserte Signal in den Zeitbereich transformiert, gefenstert und mit dem *overlapp add*-Verfahren (OLA-Verfahren) bearbeitet. Somit ist das Funkgeräte-Signal $r_\mathrm{e}(n)$ vom störenden Rauschen bereinigt, ggf. verstärkt und entzerrt.

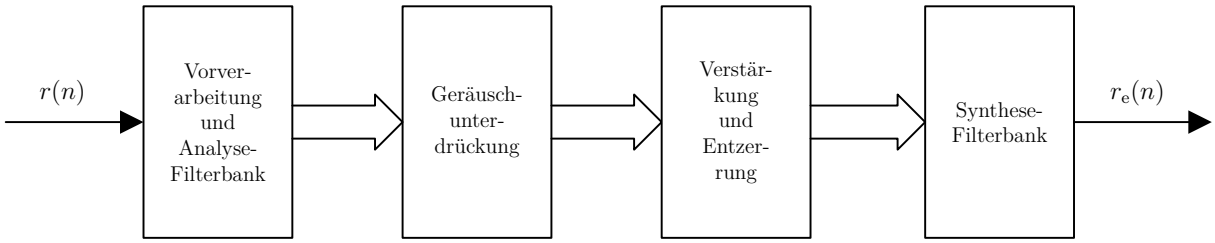


Abbildung 3.2: Funkgeräte-Signalverbesserung.

3.2 Mischung und Verstärkung der Signale

Die Mischung und Verstärkung wird hinter den Verbesserungen der Eingangssignale angewendet (siehe 3.1). Hierbei werden die Eingangssignale ggf. verstärkt und auf den gewünschten Ausgang gemischt. Die Ausgangskanäle sind hierbei der VA, die Ohr-Lautsprecher, der Funkgeräteausgang und der Team-Funk-Ausgang. Die Verstärkung der Kanäle wird so ausgelegt, dass die Eingangsdynamik gut ausgesteuert wird und es zu keinen Übersteuerungen kommt. Das Signal des Mikrofons wird mittels eines analogen Vorverstärkers ausgesteuert und benötigt somit keine digitale Verstärkung. Das Signal des Team-Funk-Eingangs wird nicht verstärkt, da dieses das Team-Funk-Ausgangssignal einer anderen Kommunikationseinheit ist und diese Signale bereits gut ausgesteuert sind. Das Signal des Funkgerätes wird ebenfalls nicht ausgesteuert, da der Nutzer beim Funkgerät einen Lautstärkeregler hat.

Bei der Mischung der Signale wird beachtet, dass pro Summierung eines zusätzlichen Kanals der Ausgang um 6 dB gemindert wird, so dass keine Übersteuerungen auftreten. Zusätzlich sind bei der Mischung noch Priorisierungen zu beachten, so dass der Funkgeräteeingang auf den Ohrhörern immer die oberste Priorität hat, da dieses beispielsweise die Kommunikationsschnittstelle zum Einsatzleiter ist. Somit wird bei einem aktiven Funkspruch auf den Ohrhörern keine Mischung der Signale des Mikrofons und des Team-Funk-Eingangs zu hören sein. Wenn der Funkgeräteeingang nicht aktiv ist, werden die anderen aktiven Signale zusammen gemischt und auf den Ohrhörern ausgegeben. Falls es erwünscht ist, kann das Funkgeräteeingangssignal auf den Team-Funk-Ausgang gemischt werden, wodurch alle Teilnehmer im Team-Funk das Funkgerätesignal hören. Dadurch können alle Teammitglieder die Ansagen des Einsatzleiters hören, ohne ein eigenes Funkgerät zu tragen. Wenn die Option aktiviert wird, ist hier ebenfalls die Priorität auf das Signal des Funkgeräts gelegt, so dass das Mikrofonsignal nicht auf den Team-Funk-Ausgang gemischt wird. Die Untersuchung wie sich das Dämpfungsverhalten der Maske in Bezug auf die Veränderung der Sprache und den Stress auswirkt, wurde auf der DAGA 2017 veröffentlicht [Mar+17]. In dieser Untersuchung ist in Bezug auf die Maske die Steigerung der Grundfrequenz und die Ausprägung des Lombard-Effekts [Lan97, S. 264 ff] aufgefallen. Damit der Nutzer es im Einsatz möglichst angenehm hat, wird das Mikrofonsignal auf die Ohrhörer gemischt, so dass der Träger eine bessere Eigenwahrnehmung hat und die Dämpfung der Maske ausgeglichen wird. Zusätzlich wird das Mikrofonsignal auf den VA, den Funkgeräteausgang und auf den Team-Funk-Ausgang gemischt. Das Team-Funk-Eingangssignal wird auf den Ohrhörern ausgegeben. Das Signal des Funkgeräteeingangs wird auf die Ohrhörer und ggf. auf den Team-Funk-Ausgang gemischt. Es sind alle Konfigurationen denkbar, so dass beispielsweise der Team-Funk auf dem linken Ohrhörer und der Funkgeräteeingang auf dem rechten Ohrhörer ausgegeben werden kann. Allerdings ist die Verständlichkeit am einfachsten, wenn alle Signale auf beiden Ohrhörern ausgeben werden, welches der typische Anwendungsfall ist.

Nach der Mischung der Signale ist noch eine Verstärkungsstufe für die Ohrhörer vorhanden, wobei diese durch den Lautstärkeregler an der Kommunikationseinheit bedient wird. Somit kann sich der Anwender die Lautstärke an den Ohrhörern einstellen. Daraufhin werden die gemischten Signale des VA's, des Funkgeräts, der Ohrhörer und des Team-Funks zur Nachverarbeitung weitergeleitet.

Kapitel 4

Mikrofon-Signalverbesserung

Die Mikrofon-Signalverbesserung (siehe Abb. 4.1) dient zur Bearbeitung des Mikrofonsignals $x(n)$ mit dem diskreten Zeitindex n, welches erst durch die Vorverarbeitung bearbeitet wird, in welcher ein Hochpass-Filter mit einer 3 dB-Grenzfrequenz von 200 Hz eingesetzt wird. Mit diesem Hochpass werden tieffrequente Anteile gefiltert, damit die Hintergrundgeräusche in dem Bereich gedämpft werden (siehe Kapitel 4.1). Daraufhin wird das vorverarbeitete Signal blockweise mit einem Hann-Fenster gewichtet, um den Leck-Effekt zu verringern und anschließend wird das Signal mit der Analyse-Filterbank in den Frequenzbereich transformiert (siehe Kapitel 4.1.2). Die Transformation wird mittels der FFT vollzogen und das spektrale Signal $X(\mu, k)$ generiert, welches von dem Frequenzband μ und dem Rahmenindex k abhängt.
Der zweite Eingang in die Mikrofon-Signalverbesserung ist das Signal des VA-Lautsprechers $y(n)$, welches zur Kompensation der Rückkopplung benötigt wird. Dafür wird $y(n)$ genau wie das Mikrofonsignal vorverarbeitet und zusätzlich um die Kopplungsdauer d von dem Lautsprecher in das Mikrofon verzögert, damit die Kopplung des Lautsprechersignals auf dem Mikrofonsignal und die berechnete Kopplung zeitlich korrelieren und die Kompensation möglich ist. Daraufhin wird die Transformation in den Frequenzbereich vollzogen und das erhaltene spektrale Signal ist $Y_{\mathrm{c}}(\mu, k - d)$. Dieses spektrale Signal wird in der Rückkopplungskompensation (siehe Kapitel 4.3) und in der Rückkopplungsunterdrückung

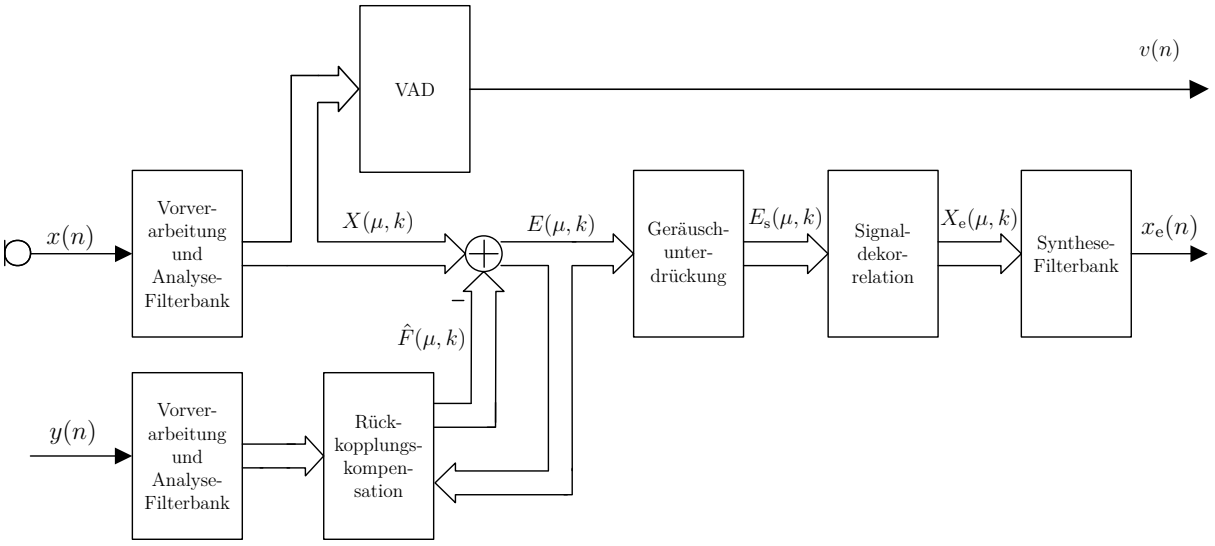


Abbildung 4.1: Mikrofon-Signalverbesserung.

verwendet. In der Rückkopplungskompensation wird die Übertragungsfunktion $\hat{H}(\mu, k)$ mit dem *Normalized Least Mean Square* (NLMS)-Algorithmus geschätzt und damit die Rückkopplung $\hat{F}(\mu, k)$ geschätzt. Die geschätzte Rückkopplung wird von dem Mikrofonsignal spektral abgezogen und das Mikrofonsignal mit geminderter Rückkopplung $E(\mu, k)$ erhalten. Dieses Signal wird in der Geräusch- und Rückkopplungsunterdrückung (siehe Kapitel 4.4) weiter verarbeitet, wobei anhand $E(\mu, k)$ des Hintergrundgeräusches und mittels $Y_{\mathrm{c}}(\mu, k-d)$ die restliche Rückkopplung geschätzt wird. Die beiden Schätzungen werden daraufhin mit einem Wiener-Filter unterdrückt. Somit sollte das Signal $E_{\mathrm{s}}(\mu, k)$ im Idealfall keine Hintergrundgeräusche und Rückkopplungen mehr enthalten. $E_{\mathrm{s}}(\mu, k)$ wird in der Dekorrelation (siehe Kapitel 4.3.4) verarbeitet. Dort wird das Signal mittels eines Frequenzversatzes dekorreliert, damit in der Rückkopplungskompensation ein möglichst großer Verstärkungsgewinn erzielt werden kann. Das verbesserte spektrale Mikrofonsignal $X_{\mathrm{e}}(\mu, k)$ nach der Dekorrelation wird mit der Synthese-Filterbank in den Zeitbereich zu dem Signal $x_{\mathrm{e}}(n)$ transformiert. In der Synthese-Filterbank wird das Signal mittels einer FFT transformiert, gefenstert und mit dem *overlap add*Verfahren zusammengefügt (siehe Kapitel 4.1.2) und das Signal $x_{\mathrm{e}}(n)$ wird erhalten.
Parallel zum verbesserten Mikrofonsignal wird die Sprachaktivität durch eine Sprachaktivitätsdetektion anhand des spektralen Mikrofonsignals $X(\mu, k)$ bestimmt (siehe Kapitel 4.2). Diese Detektion wird mittels einer Merkmalsextraktion und eines darauf folgenden Mustererkenners bestimmt. Das Signal $v(n)$ beinhaltet die Information der Sprachaktivität und wird an die Mischung und Verstärkung weiter gegeben.

4.1 Vorverarbeitung

Die Vorverarbeitung wird beim Mikrofon- und Funkgeräteeingangssignal angewendet, wobei die Vorverarbeitung einen Hochpassfilter und eine Analyse-Filterbank beinhaltet. Dabei wird die Hochpassfilterung zur Dämpfung von tieffrequenten Signalanteilen genutzt, welche beispielsweise durch ein Feuer oder durch eine Schutzbelüftungsanlage entstehen können. Nach der Hochpassfilterung wird eine Transformation vom Zeitbereich in den Frequenzbereich durch eine Analyse-Filterbank durchgeführt.

4.1.1 Hochpassfilterung

Die Hochpassfilterung wird für die Dämpfung von tieffrequenten Störsignalen und zur Entfernung des Gleichanteils verwendet. Die Pitch-Frequenz von männlichen Stimmen beginnt bei ca. 100 Hz und die von weiblichen beginnt bei ca. 200 Hz [Rap+07, S. 49], weshalb die Hochpassfilterung keine signifikante Dämpfung über 100 Hz verursachen sollte. Unter 100 Hz sind oft Störsignale mit sehr großer Energie vorhanden, wie beispielsweise der Lärm eines Feuers oder eines C-Strahlrohrs. Diese Signale können somit unterhalb 100 Hz gut gedämpft werden. Für die Filterung werden die digitalen Filter erster Ordnung aus dem verwendeten Audiocodec genutzt, welche die Übertragungsfunktion

$$H(z) = \frac{N_0 + N_1 \cdot z^{-1}}{1 - D_1 \cdot z^{-1}} \tag{4.1}$$

besitzen [TFC+16, S.25]. Dieses ist ein Filter mit unendlicher Impulsantwort (IIR-Filter) erster Ordnung in 16-Bit-Festkomma-Darstellung. Dabei sind N_0, N_1 und D_1 program-

mierbare 16-Bit Koeffizienten.
Die 3-dB-Grenzfrequenz wird durch einen Registereintrag gesetzt und kann wahlweise auf $0,0045 \cdot ADCf_\mathrm{s}$, $0,0125 \cdot ADCf_\mathrm{s}$ oder $0,025 \cdot ADCf_\mathrm{s}$ gesetzt werden, wobei die Abtastrate der Kommunikationseinheit 16 kHz beträgt. Für diese Anwendung wird die 3-dB-Grenzfrequenz auf 200 Hz gesetzt, da bei der niedrigst möglichen Grenzfrequenz von 72 Hz die Störgeräusche im Tieffrequenten durch das Filter erster Ordnung nicht stark genug gedämpft werden. Somit wird die Grenzfrequenz von 200 Hz verwendet, bei welcher die Pitch-Frequenz von tiefen männlichen Stimmen zwar leicht gedämpft werden, aber die tieffrequenten Signalanteile und der Gleichanteil stärker gedämpft werden.

4.1.2 Analyse- und Synthese-Filterbank

Die Teilbandverarbeitung ist in der Signalverarbeitung ein zentrales Element zur Zeit-Frequenz-Analyse, so dass der zeitliche und spektrale Aspekt zusammen untersucht werden kann. Im Gegensatz zu der Aufwandsreduktion und der Analyse von zeitlichen und spektralen Aspekten steht die Laufzeit. Durch die Transformation durch eine Analyse-Filterbank in den Frequenzbereich und durch die Rücktransformation durch eine Synthese-Filterbank in den Zeitbereich wird das Signal verzögert. Diese Verzögerung kann im Zusammenhang mit Echoeffekten und zeitkritischen Frequenzbereichsalgorithmen störend sein. Allerdings sind Algorithmen zur Rückkopplungskompensation, Störgeräuschreduktion usw. effizienter [Hän+04]. Die Signalverzögerung wird hierbei als Kompromiss für die zeitliche und spektrale Analyse und für die Aufwandsreduktion gesehen und die Filterbänke werden so ausgelegt, dass mit der Signalverzögerung kein Echoeffekt bei der Wiedergabe entsteht.

4.1.3 Kurzzeit-Fouriertransformation der Filterbänke

Die Transformation vom Zeit- in den Frequenzbereich kann durch eine Diskrete Fourier-Transformation (DFT) durchgeführt werden [Neu12]. Bei dieser Transformation wird das diskrete Eingangssignal $x(n)$ mit dem diskreten Zeitindex n mit einem Fenster h_ana multipliziert. Dieses Tiefpassfilter wird als Analyse-Fenster bezeichnet. Die Berechnung des Kurzzeitspektrums mittels der DFT mit dem Analyse-Fenster und dem Eingangssignal ergibt sich zu

$$X(\mu,k) = \sum_{n=0}^{N_\mathrm{DFT}-1} h_\mathrm{ana}(n)x(n+kR)e^{-j\frac{2\pi}{N_\mathrm{DFT}}\mu n}, \tag{4.2}$$

wobei $X(\mu,k)$ das erhaltene Kurzzeitspektrum mit dem Frequenzindex μ, dem Rahmenindex k und der DFT-Ordnung N_DFT ist. Der Rahmenindex wird durch den Rahmenversatz, bei dem R Abtastwerte zu einen Rahmen zusammengefasst werden, und den diskreten Zeitindex durch den Zusammenhang $k = \lfloor n/R \rfloor$ berechnet.

Mit der Synthese-Filterbank werden die Signale vom Frequenzbereich in den Zeitbereich transformiert, welches durch die inverse diskrete Fourier-Transformation

$$x_k(n) = \begin{cases} \mathrm{IDFT}\{X_e(\mu,k)\} \ , & \text{wenn } n = 0,1,\ldots,N_\mathrm{DFT}-1 \\ 0, & \text{sonst} \end{cases} \tag{4.3}$$

geschieht. Die resultierenden Signalblöcke im Zeitbereich haben jeweils die Länge N_{DFT} und sollen überlappend addiert werden. Die Signalblöcke werden mit einem Synthesefenster h_{syn} wie folgt gewichtet:

$$x_e(n) = \sum_{k=-\infty}^{\infty} h_{\mathrm{syn}}(n - k \cdot R) \cdot x_k(n - k \cdot R). \tag{4.4}$$

Durch die Überlappung wird das Signal um N_{DFT}/R Rahmen verzögert, wobei N_{DFT} die Länge des Analysefensters ist. Daraufhin wird die überlappende Additions-Methode angewendet [Bos+13, S. 246 ff][Rao+11, S. 27 ff], wobei sich die gefensterten Signalblöcke überlappen und zur Erzeugung des Zeitbereichssignals addiert werden.

Design des Analyse- und Synthesefensters

Das Design des Analyse- und Synthesefensters ist sehr wichtig, damit eine perfekte Rekonstruktion gegeben wird, der Leck-Effekt (*Leakage*-Effekt) minimiert wird und die Spiegelfrequenzen unterdrückt werden [Hav+08][Mey14]. Die Rekonstruktion wird durch die verbundene Betrachtung des Analyse- und des Synthesefensters realisiert, wobei eine FFT-Ordnung von $N_{\mathrm{DFT}} = 128$ und ein Rahmenversatz von $R = 32$ im Hann-Fenster zur Gewichtung gewählt werden. Nach der Fensterung wird eine FFT durchgeführt [Mer12], wobei diese im Vergleich zu einer DFT deutlich effizienter ist; wobei die DFT einen Aufwand von N_{DFT}^2 und die FFT einen Aufwand von nur $N_{\mathrm{DFT}} \cdot \log(N_{\mathrm{DFT}})$ aufweist. Bei dem genutzten C5515-DSP ist ein Co-Prozessor zur effektiven Berechnung der FFT verbaut, welcher wiederum deutlich weniger Rechenzyklen braucht, als eine klassische Implementierung. Dieser Unterschied ist in der Tabelle 2.1 dargestellt, wobei der Koprozessor bei der FFT-Ordnung von 128 dreimal so schnell ist wie eine auf dem Hauptprozessor implementierte FFT [C55a]. Somit ist eine Verwendung der FFT auf dem Prozessor sehr effektiv. Die Ordnung der FFT wird durch N_{fft} ausgedrückt, wobei zur Berechnung innerhalb der Algorithmen lediglich $N_{\mathrm{fft}}/2 + 1$ Stützstellen auf Grund der Symmetrie benötigt werden. Die erste Stützstelle stellt den Gleichanteil des Signals dar [Kam+09, S. 268 ff].

4.2 Sprachaktivitätserkennung

Die Sprachaktivitätserkennung der Kommunikationseinheit wurde 2015 auf der DAGA veröffentlicht [Bro+15]. Eine VAD wird in verschiedenen Bereichen für diverse Anwendungen genutzt und ist dort ein notwendiger Bestandteil für die Funktion [Gra+15]. Anwendungen können beispielsweise Innenraumkommunikationssysteme, Konferenzsysteme und Freisprecheinrichtungen sein. Bei diesen Anwendungen wird die VAD oft durch das SNR auf Basis des Betrags des geschätzten Hintergrundgeräuschs und des Betrags des Gesamtsignals entschieden [Lük+11]. Wenn hierbei das SNR groß genug ist, wird der betrachtete Bereich als Sprache erkannt. Ein solches Verfahren funktioniert allerdings nur bei einer sehr guten Geräuschschätzung; diese wird typischerweise durch eine geeignete Signalglättung umgesetzt. Diese VAD ist sehr empfindlich gegenüber Hintergrundgeräuschen, welche instationär sind und ihren Pegel sehr schnell ändern. Bei der schnellen Änderung des Störsignals können Geräuschschätzungen, welche durch eine Signalglättung umgesetzt sind, diesem Verlauf nicht mehr folgen und somit wird das Hintergrundgeräusch

sehr wahrscheinlich falsch geschätzt, wodurch bei der VAD Fehlerkennungen möglich sind. Dabei können instationäre Geräusche mit einem Pegel über dem stationären Hintergrundgeräusch potentiell als Sprache fehlerkannt werden.

4.2.1 Problemstellung

Bei der Kommunikationseinheit sollen wie in Kapitel 2.2 mit der Sprachaktivitätserkennung die Atem- und Hintergrundgeräusche gefiltert werden, und der Funkkanal für den Team-Funk alloziert werden. Die Funktionalität der VAD ist für die Funkkanal-Allozierung essenziell. Es gibt 2 Funkkanäle und es können bis zu 10 Teilnehmer gleichzeitig aktiv sein. Wenn eine hohe Fehlerkennungsrate der VAD vorhanden ist, ist eine Kommunikation über den Team-Funk nicht mehr möglich. Daher muss die VAD der Kommunikationseinheit sehr zuverlässig sein und eine hohe Erkennungsrate haben. VAD's auf Basis von Signalglättung funktionieren für die Kommunikationseinheiten von Atemvollschutzmasken allerdings nicht. Der Pressluftatmer (SCBA, *Self-contained breathing apparatus*) erzeugt laute Atemgeräusche, welche in Verbindung mit der Kommunikationseinheit verstärkt über die Lautsprecher ausgegeben werden.

Bei der Funkkommunikation sind die Atemgeräusche ebenfalls sehr störend, da der Leistungspegel von Sprache zu Atemgeräusch stark schwankt und dadurch die Lautstärkeeinstellung beispielsweise bei der Leitzentrale auf einem Headset sehr schwierig ist. Die Sprachverständlichkeit sinkt aufgrund dieser Problematik ebenfalls.

Um die Verständlichkeit zu steigern und eine Team-Funk-Kommunikation zu ermöglichen, werden mit der Sprachaktivitätsdetektion die Atemgeräusche mit einem Mustererkenner erkannt und herausgefiltert. Dazu werden auf Basis des Mikrofonspektrums Merkmale extrahiert, welche dem Mustererkenner zur Verfügung gestellt werden. Ziel des Mustererkenners ist die Unterscheidung der Klassen Nebensprecher, Pausen, Ausatmen, Einatmen und Sprache, wobei die verschiedenen Klassen in der Abb. 4.2 dargestellt sind. Die Klassen unterscheiden sich spektral sehr stark voneinander: Die Pause besitzt die geringste Energie und beim Ausatmen ist die Energie über die Frequenz ungefähr gleich verteilt, ebenso beim Einatmen, wobei hier die Energie wesentlich höher ist; die Sprache unterscheidet sich aufgrund der variierenden Spektralverteilung deutlich von den anderen Klassen. Dabei ist meist tieffrequent die höchste und zu den höheren Frequenzen immer weniger Energie vorhanden. Die nicht dargestellte Klasse sind die Nebensprecher. Diese ähnelt der Sprachklasse, da diese lediglich ein anderes Dämpfungsverhalten als die Sprache der die Maske tragenden Person aufweist.

Beim Mustererkenner werden ein Codebuch und ein neuronales Netz hinsichtlich der Erkennungsraten, Rechenoperationen und Speicherbedarf verglichen. Außerdem wird zur Steigerung der Erkennungsrate eine lineare Diskriminanzanalyse zwischen der Merkmalsextraktion und dem Mustererkenner vorgestellt.

Bei der Sprachaktivitätsdetektion ist der Unterschied zwischen dem Ausatmen und den Zischlauten von besonderer Bedeutung, da es beispielsweise sehr wichtig ist, ob „*person*“ oder „*persons*“ im Englischen gesprochen wird, da der Einsatzleiter darauf unterschiedlich reagieren muss. Die Schwierigkeit besteht in der Unterscheidung durch die spektrale Ähnlichkeit. Einerseits dürfen keine Sprachpassagen unterdrückt werden und andererseits darf kein Ausatmen fehlerkannt werden, wodurch der Funkkanal für den Team-Funk allo-

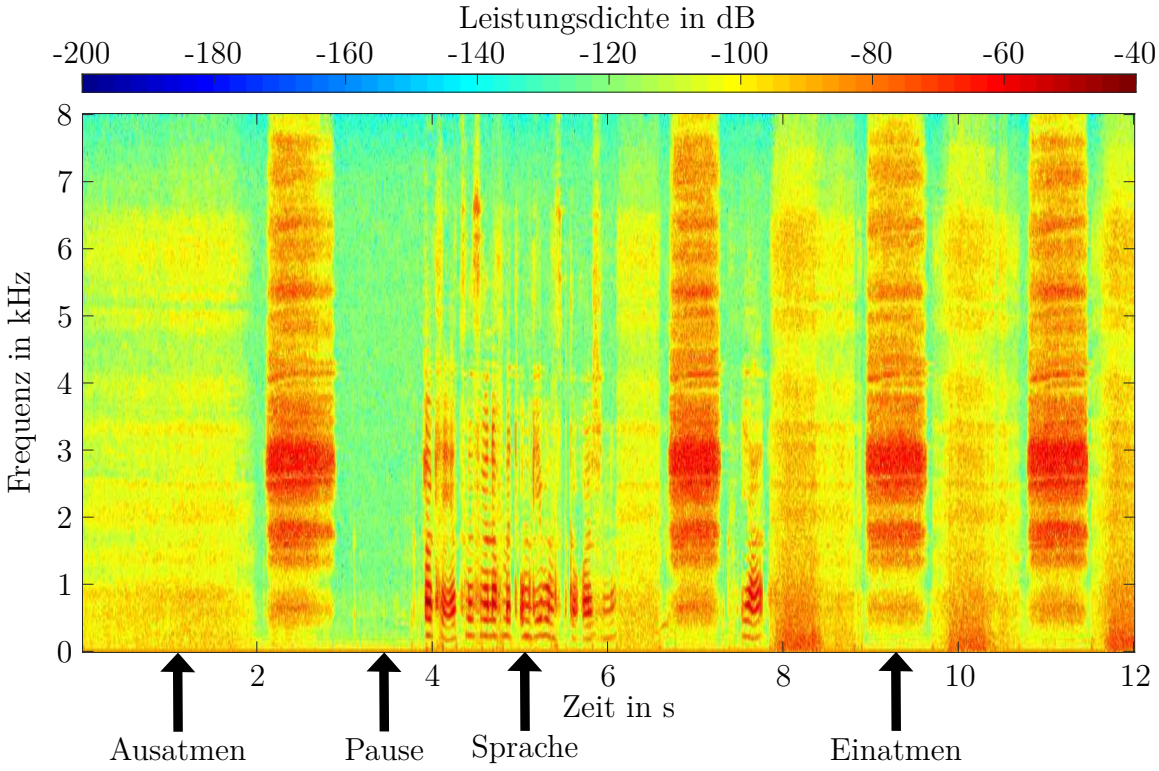


Abbildung 4.2: Spektrogramm vom Mikrofonsignal einer Kommunikationseinheit.

ziert werden würde. Wenn diese Fehlallozierung auftritt, kann es bei mehreren Personen dazu führen, dass die zwei Sprachkanäle dauerhaft besetzt sind und dadurch keine Funkkommunikation mehr möglich ist. Um die Fehlerkennungen zu minimieren, wird nach dem Mustererkenner eine Nachverarbeitung durchgeführt. Diese beinhaltet eine Klassenabhängige Glättung des Ausgangs des Mustererkenners, so dass die Zustände nicht schnell wechseln und die Fehlerkennungen reduziert werden. Die gesamte Sprachaktivitätsdetektion ist in der Abb.4.3 dargestellt.

Zusätzlich muss bei einer VAD das sogenannte *Front-End-Speech-Clipping* [Dav+00, S. 218 f] betrachtet werden, welches bei zu später Detektion der Sprache den Anfang abschneidet. Die VAD hat durch den Rahmenversatz, die Filterbank und Glättung in der Nachverarbeitung eine Laufzeit. Diese muss so gering wie möglich gehalten werden und liegt in Summe bei 12 ms, so dass der Effekt kaum bemerkbar ist.

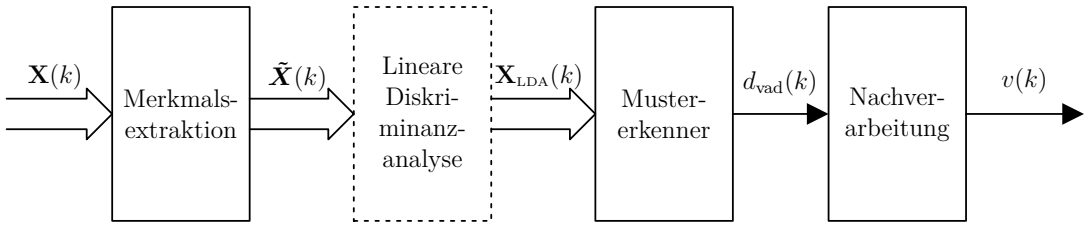


Abbildung 4.3: Signalflussgraph der Sprachaktivitätserkennung.

Die lineare Diskriminanzanalyse kann als optional angesehen werden, weshalb sie gestrichelt dargestellt ist. Wenn die Anwendung der linearen Diskriminanzanalyse einen

Erkennungsgewinn erzielen kann und die verwendeten Rechenoperationen im Verhältnis zu dem Erkennungsgewinn zu rechtfertigen sind, ist der Signalfluss wie in Abb. 4.3. Andernfalls sind die extrahierten Merkmale $\tilde{\boldsymbol{X}}(k)$ das Eingangssignal des Mustererkenners. Dabei sind die Elemente des Eingangsvektors

$$\boldsymbol{X}(k) = [X(0,k), \cdots, X(N_{\text{fft}}/2-1,k)]^{\text{T}} \tag{4.5}$$

und die Elemente des Merkmalsvektors

$$\tilde{\boldsymbol{X}}(k) = [X_{\text{feat}}(0,k), \cdots, X_{\text{feat}}(N_{\text{m}}-1,k)]^{\text{T}}, \tag{4.6}$$

wobei die Anzahl der Merkmale N_{m} ist. Nach der linearen Diskriminanzanalyse kann die Anzahl der Merkmale ggf. noch reduziert worden sein.

4.2.2 Merkmalsextraktion

Ein Mustererkenner nutzt Signaleigenschaften für die Klassifikation aus. Daher wird das Mikrofonsignal zuerst einer Merkmalsextraktion unterzogen. Diese verarbeitet das Eingangssignal so, dass die relevanten Merkmale herausgearbeitet werden können. Ziel ist es, dass die Eigenschaften des Eingangssignals durch möglichst wenige signifikante Merkmale beschrieben werden, um den Rechenaufwand zu minimieren. Der Eingangsvektor des Mikrofonspektrums ist $X(\mu,k)$, welcher in der Analyse-Filterbank 4.1.2 erzeugt wird. In der Abb. 4.4 ist der Signalflussgraph der Merkmalsextraktion dargestellt.

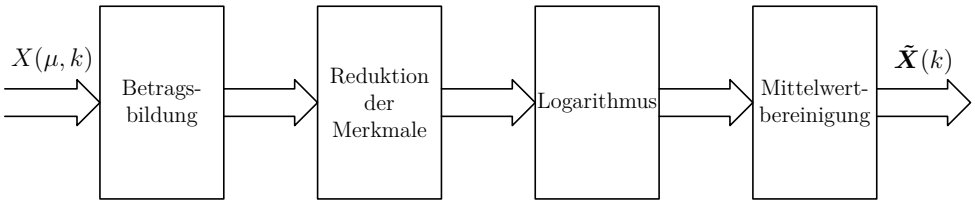


Abbildung 4.4: Signalflussgraph der Merkmalsextraktion.

Von dem Mikrofonspektrum wird eine aufwandsreduzierte genäherte Betragsbildung

$$\Big|X(\mu,k)\Big| \approx \Big|\Re\Big\{X(\mu,k)\Big\}\Big| + \Big|\Im\Big\{X(\mu,k)\Big\}\Big| \tag{4.7}$$

berechnet. Im Allgemeinen ist die Berechnung einer komplexen Betragsbildung

$$|X(\mu,k)| = \sqrt{\Re\{X(\mu,k)\}^2 + \Im\{X(\mu,k)\}^2}, \tag{4.8}$$

wobei der Aufwand durch die quadratische Wurzel und die Quadrierung des Real- und Imaginärteils hoch ist. Die Phaseninformation ist für die Unterscheidung zwischen den Klassen nicht notwendig und wird beispielsweise bei Mel-Frequenz-Cepstrum-Koeffizienten (MFCC) Berechnungen nicht ausgewertet [Tan+08]. Für die Merkmalsextraktion ist der exakte Wert der Betragsbildung nicht notwendig. Eine Betragsschätzung mit reduziertem Aufwand ist daher ausreichend. Auf Basis dieses geschätztem Betragsspektrums wird die Anzahl der Merkmale in der genäherte Mel-Filterung reduziert.

Genäherte Mel-Filterung

Das menschliche Gehör hat ein unterschiedlich ausgeprägtes Hörverhalten in Bezug auf den Frequenzbereich. Die Frequenzen können dabei in Bereiche zusammengefasst werden, wobei im Hochfrequenten jeder Bereich mehr Frequenzen beinhaltet als im Tieffrequenten. Somit ist das Hörvermögen bei tiefen Frequenzen besser aufgelöst als bei hohen Frequenzen. Dieses Phänomen wurde bereits 1937 von Stevens, Newman und Volkmann [Ste+37] untersucht. Daher wird die Reduktion der Merkmale durch eine genäherte Mel-Filterung [Wen04, S.47 ff] umgesetzt. Der nichtlineare Zusammenhang zwischen der Frequenz- und der Mel-Skala kann durch die Gleichung

$$\text{mel}(f) = 2595 \cdot \log_{10}\left(1 + \frac{f}{700}\right) \tag{4.9}$$

ausgedrückt werden, wobei f die lineare Frequenz in Hz ist.
Dieser Zusammenhang wird bei der Mel-Filterung typischerweise durch Dreiecke ausgedrückt, bei welchen die Breite zu höheren Frequenzen zunimmt, um das zuvor erwähnte Auflösungsvermögen widerzuspiegeln [Hol12]. Die Frequenzstützstellen in jedem Dreieck werden mit der Amplitude des Dreiecks gewichtet und daraufhin aufsummiert, wodurch bei jedem Dreieck eine Mel-Stützstelle entsteht. Durch die Gewichtung mit der Amplitude sind die Mel-Stützstellen normiert. Diese Mel-Filterung ist in der Abb. 4.5 dargestellt.
Bei der Signalverarbeitung in der Kommunikationseinheit wird diese Mel-Filterung nur angenähert. Der verwendete Ansatz ist in Abb. 4.6 dargestellt. Die Dreiecke sind durch Rechtecke repräsentiert. Die Anzahl der Stützstellen pro Rechteck sind so gewählt, dass die Mittlung durch einen Bitshift realisiert werden kann, wodurch die Rechenleistung im Gegensatz zu einer herkömmlichen Division minimiert wird. Die Auflösung ist sehr ähnlich zu der Mel-Skala und ist für den Mustererkenner ausreichend, da somit die signifikanten Bereiche der einzelnen Klassen gut dargestellt sind. Die Sprache hat im Tieffrequenten

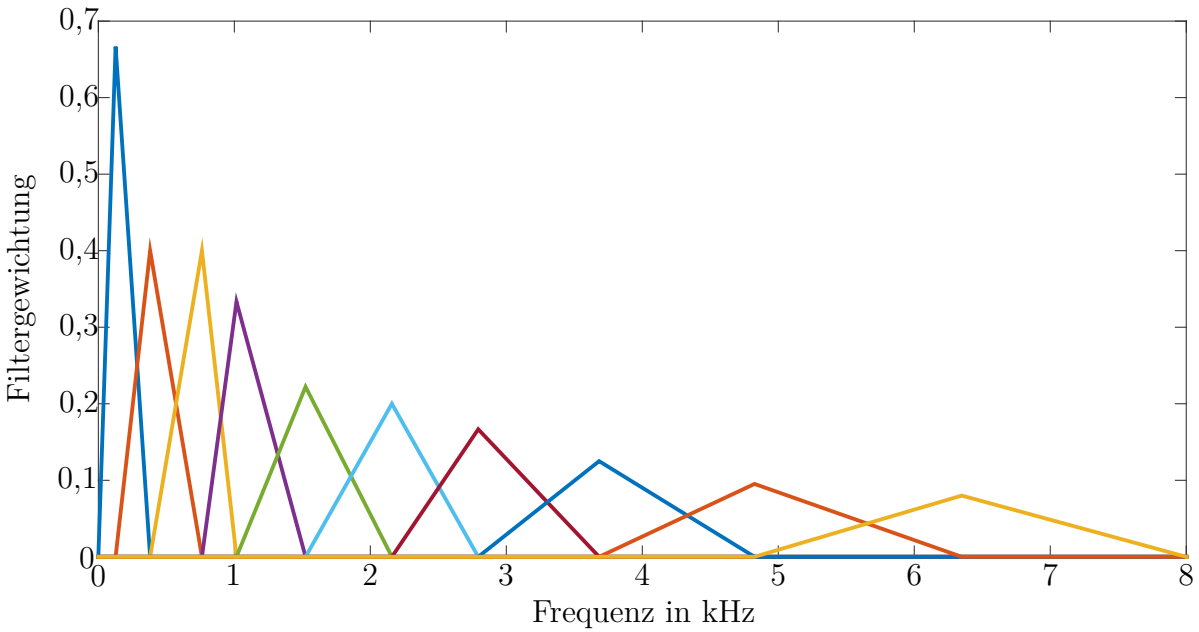


Abbildung 4.5: Mel-Filterung mit 10 Stützstellen.

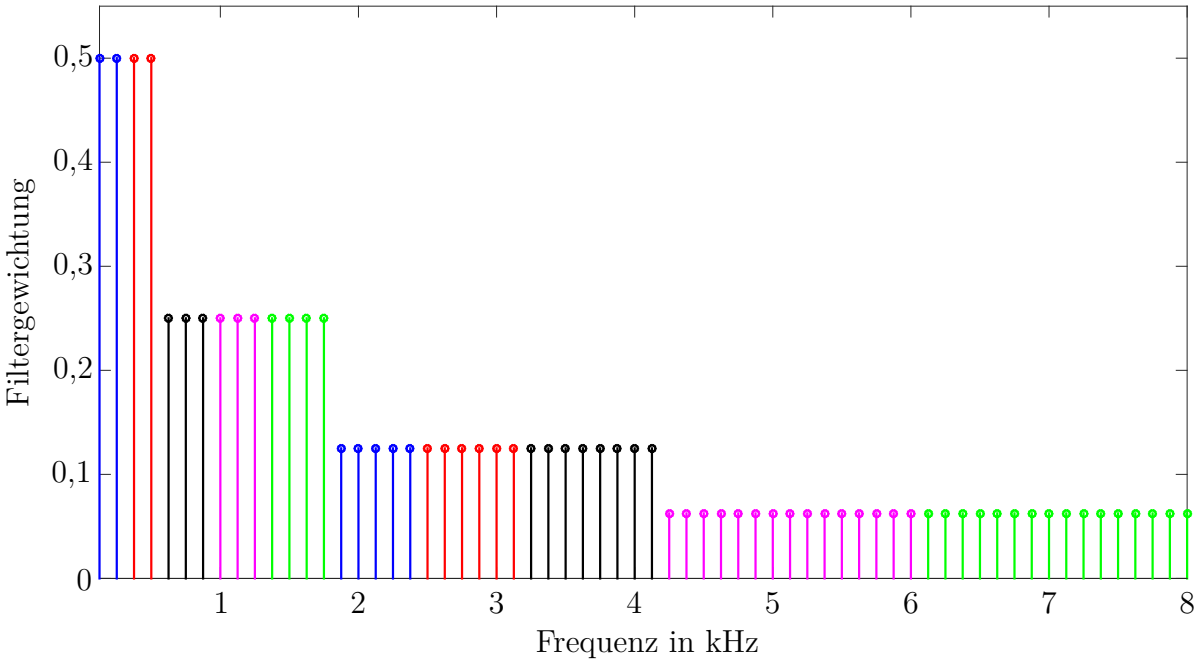


Abbildung 4.6: Genäherte Mel-Filterung mit 10 Stützstellen.

mehr Harmonische und dadurch muss die Auflösung hier höher sein als im Hochfrequenten. Die Klassen Ein- und Ausatmen haben bis 3,5 kHz unterschiedliche Leistungsdichten und somit ist hier der interessante Bereich ebenfalls im Tieffrequenten. Die Pause ist breitbandig ohne Energie und daher ist hier die Auflösung nicht von Bedeutung. Jedes Rechteck, gekennzeichnet in einer einheitlichen Farbe, stellt eine Mel-Stützstelle dar und der Mel-Index ist m. Insgesamt werden die $N_{\mathrm{Sub}} = 65$ Frequenzstützstellen in der Merkmalsextraktion auf $N_{\mathrm{M}} = 10$ Mel-Stützstellen reduziert. Somit ist das frequenzselektive Wahrnehmungsvermögen des menschlichen Gehörs angenähert und der Rechenaufwand für die nachfolgenden Schritte minimiert.

Logarithmierung

Die Merkmalsextraktion ist mittels der Mel-Filterbank an das frequenzselektiven Verhalten des menschlichen Gehörs angelehnt. Dabei ist die Betrachtung der Lautheitsempfindung nicht eingebunden. Dieses ist jedoch notwendig, damit der Mustererkenner bestmöglich zwischen verschiedenen Zuständen unterscheiden kann. Die Lautstärke ist dabei das Maß für die Lautheitsempfindung des menschlichen Gehörs [Her+13; Zwi82]. Diese Lautstärke L_{s} ist bei 1 kHz gleich dem Schalldruckpegel und wird in phon gemessen:

$$L_{\mathrm{s}}(1\,\mathrm{kHz}) = 20 \cdot \log_{10} \frac{p_{\mathrm{eff}}}{p_{\mathrm{eff},0}} \mathrm{phon} \tag{4.10}$$

Dabei ist p_{eff} der effektive Schalldruck und der Bezugsschalldruck $p_{\mathrm{eff},0} = 2 \cdot 10^{-5}\,\mathrm{Pa}$. Die untere Wahrnehmungsschwelle des Gehörs liegt bei ca. $L_{\mathrm{s}} = 0\,\mathrm{Phon}$ und die Schmerzgrenze liegt bei ca. $L_{\mathrm{s}} = 120\,\mathrm{Phon}$. Das menschliche Auflösungsvermögen liegt bei ca. 1 Phon.

Das Gehör hat einen näherungsweise logarithmischen Verlauf (siehe Gl. 4.10) [Wei09,

S.60 ff] und somit ergibt sich durch

$$X_{\mathrm{dB}}(m,k) = 20 \cdot \log_{10}\Big(X_{\mathrm{M}}(m,k)\Big) \tag{4.11}$$

das logarithmierte Mel-Spektrum.

Die Logarithmierung ist auf dem DSP in einer internen Bibliothek enthalten. Die Berechnung eines Logarithmus zu Basis 10 braucht 35-Zyklen und zur Basis 2 36-Zyklen [C55b, 4-77 ff]. Allerdings wird für die Merkmalsberechnung nicht der exakte Wert des Logarithmus benötigt, daher ist eine gute Näherung ausreichend. Dies erlaubt es den Rechenaufwand zu reduzieren.

Im Folgenden wird der Logarithmus zur Basis 2 ($\log_2$) verwendet. Durch dessen Eigenschaften lässt sich eine effiziente Näherung bestimmen. Dafür wird der näherungsweise lineare Zusammenhang im Bereich $1 \leq x \leq 2$ zwischen dem Eingang x und dem $log_2(x)$ ausgenutzt, welcher in der Abb. 4.7 dargestellt ist. Zusätzlich wird bei der Näherung der

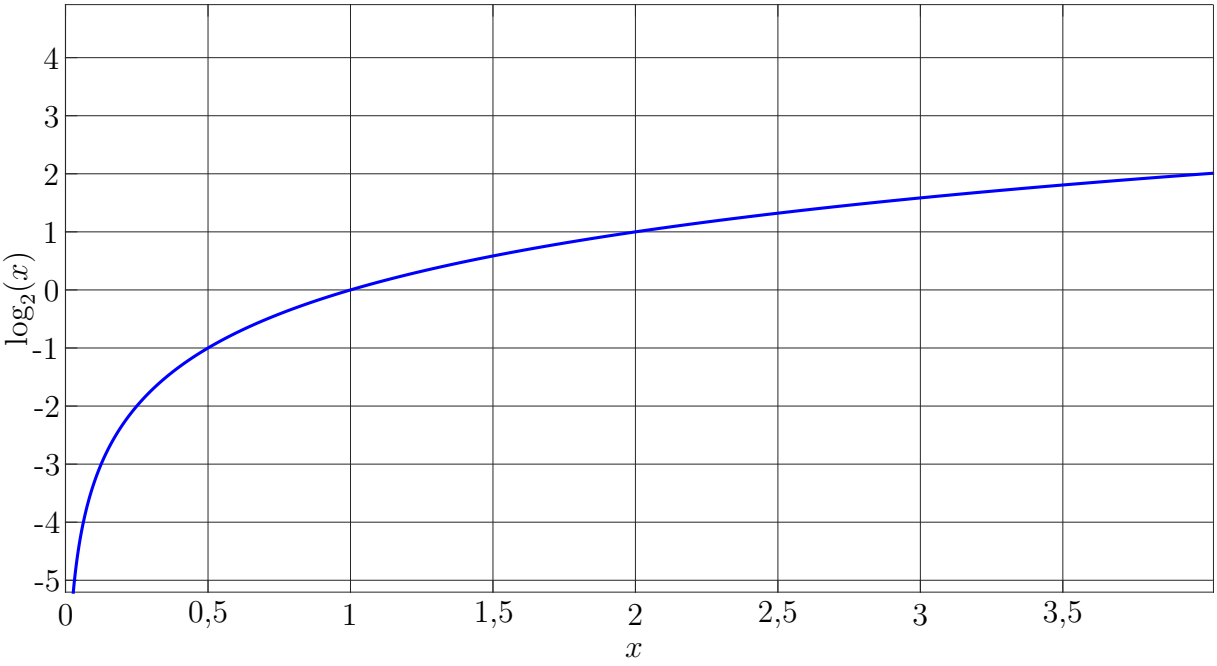


Abbildung 4.7: Logarithmus dualis zur Darstellung des linearen Zusammenhangs zwischen eins und zwei auf der x-Achse.

Zusammenhang zwischen den Werten des $\log_2$ und einem Bitshift ausgenutzt. Wenn im $\log_2$ um einen Wert inkrementiert wird, entspricht dieses im Linearem einem Bitshift. Somit muss das Ergebnis y im linearen in den Bereich $1 \leq x \leq 2$ durch Bitshifts gebracht werden, wobei die Anzahl der benötigten Shifts in N_{Shift} gespeichert wird. Somit ergibt sich bis hier folgender Zusammenhang:

$$y = \frac{x}{(2^{N_{\mathrm{Shift}}})} \tag{4.12}$$

Da für $1 \leq x \leq 2$ näherungsweise $0 \leq \log_2(x) \leq 1$ gilt, muss 1 von y subtrahiert werden. Insgesamt ergibt sich folgender Zusammenhang:

$$\log2(x) \approx y - 1 + N_{\text{Shift}}. \tag{4.13}$$

Diese Näherung und der genaue Logarithmus dualis ist in Abb. 4.8 dargestellt. Die entstandene Ungenauigkeit ist maximal 0,0864 und die Erkennungsrate des Mustererkenners ist gleich geblieben.

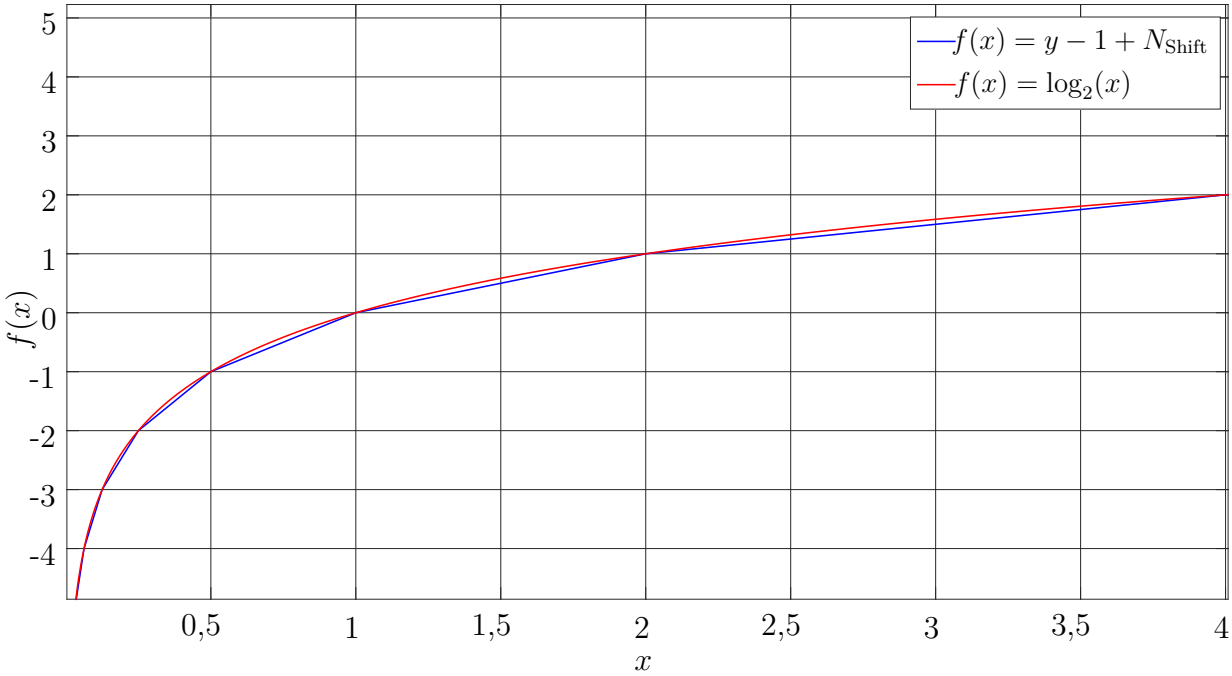


Abbildung 4.8: Näherung des Logarithmus dualis im Vergleich zum genauen Logarithmus dualis.

Der Rechenaufwand minimiert sich durch diese Näherung bei der maximalen Anzahl der Bitshifts auf maximal 18 Zyklen und minimal 5-Zyklen, wenn $1 \leq x \leq 2$ bereits gegeben ist. Somit hat sich der Rechenaufwand gegenüber der genaueren Berechnung deutlich reduziert.

Mittelwertbereinigung

Bei der Mustererkennung ist es wichtig, dass mittels der Merkmale möglichst alle Zustände abgebildet sind. Daher sollten die Merkmale so viele Informationen wie möglich enthalten. Um hierbei die Leistung nicht mitzubetrachten, werden die Merkmale vom Mittelwert befreit [Pfi+08, S. 107 ff.]. Im erstem Zuge wird der Mittelwert

$$\overline{X}_{feat}(k) = \left(\sum_{m=0}^{N_{\text{M}}-1} X_{\text{dB}}(m,k) \right) / N_{\text{M}} \tag{4.14}$$

berechnet, dieses geschieht durch die Addition von X_{dB} über die Mel-Bänder und dividiert durch die Anzahl der Mel-Bänder. In der Mittelwertbereinigung werden von X_{dB} der

Mittelwert $\overline{X}_{feat}$ subtrahiert:

$$\tilde{X}(m,k) = X_{\mathrm{dB}}(m,k) - \overline{X}_{feat}(k). \tag{4.15}$$

In dem Vektor

$$\tilde{\boldsymbol{X}}(k) = \left[\tilde{X}(0,k),\, \cdots,\, \tilde{X}(N_{\mathrm{M}}-1,k)\right]^{\mathrm{T}} \tag{4.16}$$

sind die vom Mittelwert bereinigten Merkmale gespeichert. Dieser Vektor wird dem Mustererkenner präsentiert und ggf. wird vor dem Mustererkenner eine lineare Diskriminanzanalyse zur Steigerung der Erkennungsraten auf Basis der Merkmale berechnet.

4.2.3 Lineare Diskriminanzanalyse

Die lineare Diskriminanzanalyse (LDA) dient der optimalen Trennung von Merkmalen aufgrund deren statistischen Eigenschaften (siehe [Sch07, S. 181 ff.]), wobei es sich um die Zuordnung von Merkmalen zu definierten Klassen handelt. Das Ziel der LDA ist es, eine Transformationsmatrix $\boldsymbol{\Theta}_{\mathrm{LDA}}$ zu finden, welche die Merkmale $\tilde{\boldsymbol{X}}(k)$ so transformiert, dass die Varianz innerhalb der Klassen minimiert und die zwischen den Klassen maximiert wird. Durch diese Vorgehensweise der LDA ist es für den Merkmalerkenner einfacher zwischen den Klassen zu unterscheiden. Grundsätzlich sind die Merkmale innerhalb einer Klasse ähnlicher, wodurch die Erkennungsrate steigt. Zusätzlich kann durch die LDA eine Dimensionsreduktion $N_{\mathrm{LDA}} \leq N_{\mathrm{M}}$ durchgeführt werden, welche durch die Dimensionen der Transformationsmatrix bestimmt wird. Der Merkmalsvektor $\tilde{\boldsymbol{X}}(k)$ wird mit der Transformationsmatrix multipliziert. Der Merkmalsvektor hat die Länge N_{M} und die Transformationsmatrix die Dimension $N_{\mathrm{M}} \times N_{\mathrm{LDA}}$, so dass die Anzahl der Merkmale von N_{M} auf N_{LDA} reduziert wird.
Während der Laufzeit wird die Transformationsmatrix $\boldsymbol{\Theta}_{\mathrm{LDA}}$ lediglich mit dem Merkmalsverktor $\tilde{\boldsymbol{X}}(k)$ multipliziert, wodurch sich der durch die LDA optimierte Vektor

$$\mathbf{X}_{\mathrm{LDA}}(k) = \boldsymbol{\Theta}_{\mathrm{LDA}}^{\mathrm{T}} \cdot \tilde{\boldsymbol{X}}(k) \tag{4.17}$$

ergibt. Die Transformationsmatrix wird durch die Datenbank der jeweiligen Klassen im Vorhinein bestimmt und benötigt somit keine Rechenleistung während der Laufzeit.
Für die Berechnung der Transformationsmatrix $\boldsymbol{\Theta}_{\mathrm{LDA}}$ wird eine Datenbank mit den jeweiligen Klassen Nebensprecher, Pause, Ausatmen, Einatmen und Sprache benötigt. Dafür müssen aufgenommene Dateien mit den zugehörigen Klassen markiert und zugeschnitten werden. Auf Basis dieser Datenbank werden für dass Training die Merkmalsvektoren für jede Klasse berechnet. Mit diesen Merkmalsvektoren wird die Transformationsmatrix berechnet. Dafür wird im ersten Schritt für jede der Klassen der Mittelwertvektor

$$\boldsymbol{\mu}_{\mathrm{K}}(j) = \frac{1}{N_{\mathrm{K}}(j)} \sum_{i=0}^{N_{\mathrm{K}}(j)-1} \tilde{X}(j,i) \tag{4.18}$$

über die Anzahl $N_{\mathrm{K}}(j)$ der Merkmalsvektoren der Klasse j gebildet. Mit diesem Mittelwert der jeweiligen Klasse wird die Kovarianzmatrix

$$\boldsymbol{\Sigma}_{\mathrm{K}}(j) = \frac{1}{N_{\mathrm{K}}} \sum_{i=0}^{N_{\mathrm{K}}(j)-1} \Big(\tilde{X}(j,i) - \mu_{\mathrm{K}}(j)\Big)\Big(\tilde{X}(j,i) - \mu_{\mathrm{K}}(j)\Big)^{T} \tag{4.19}$$

jeder Klasse j berechnet. Mit den Merkmalsvektoren des Trainings über alle Klassen werden der Mittelwert $\boldsymbol{\mu}$ und die Kovarianzmatrix $\boldsymbol{\Sigma}$ berechnet. Dabei ist die Gesamtzahl der Merkmalsvektoren des Trainings

$$N_{\mathrm{G}} = \sum_{j=1}^{N_{\mathrm{C}}} N_{\mathrm{K}}(j) \tag{4.20}$$

mit N_{C} Klassen gegeben.
Zur Maximierung der Varianz unter den Klassen und zur Minimierung der Varianz innerhalb der Klassen werden die N_{LDA} größten Eigenwerte von $\overline{\boldsymbol{\Sigma}}^{-1} \cdot \boldsymbol{\Sigma}_{\mathrm{K}}$ berechnet [Bis06], wobei

$$\overline{\boldsymbol{\Sigma}} = \frac{1}{N_{\mathrm{G}}} \sum_{j=0}^{N_{\mathrm{C}}} N_{\mathrm{K}}(j) \cdot \mu_{\mathrm{K}}(j) \tag{4.21}$$

ist. Aus den N_{LDA} größten Eigenwertvektoren bildet sich somit die Transformationsmatrix $\boldsymbol{\Theta}_{\mathrm{LDA}}$.

4.2.4 Mustererkennung

Die Mustererkennung ist aus der Nachahmung des menschliches Wahrnehmungsvermögen entstanden [Hof+15]. Dabei können verschiedenste Problemstellungen behandelt werden. Beispielsweise kann ein Münzautomat unterschiedliche Münzen aufgrund deren Eigenschaften erkennen. Ebenso ist die Anwendung von Mustererkennern in der Bild- und Sprachsignalverarbeitung weit verbreitet. Bei der Bildsignalverarbeitung kann dieses eine Gesichtserkennung in Kameras sein und in der Sprachsignalverarbeitung beispielsweise die Spracherkennung. Im Bereich der Kommunikationseinheiten für Atemvollschutzmasken soll die Sprachaktivität mittels verschiedener Zustände anhand von Informationen vom Mikrofon bestimmt werden. Die Zustände, auch Klassen genannt, werden mittels eines Mustererkenners auf Basis extrahierter Merkmale bestimmt. Die Zustände gliedern sich in Nebensprecher, Pause, Ausatmen, Einatmen und Sprache. Wenn der Mustererkenner die Zustände sehr gut erkennt, kann mit dieser Kenntnis nur die Sprache übertragen und alle Störgeräusche unterdrückt werden.

Die Mustererkennung kann mittels verschiedener Algorithmen umgesetzt werden, welche beispielsweise ein neuronales Netz, ein Codebuch oder ein Gauß'sches Mischmodell sein können.

Für die Mustererkennung im Bereich der Atemgeräuscherkennung werden ein neuronales Netz und ein Codebuch miteinander in Bezug auf deren Erkennungsraten, den verwendeten Rechenoperationen und des Speicherbedarfs verglichen. Andere Mustererkenner wurden nicht weiter verfolgt, da beispielsweise das Gauß'sche Mischmodell eine höhere Rechenkomplexität aufgrund der verwendeten Gaußglocken benötigt [Goo+18]. Dies steht im Widerspruch zu der erwünschten Berechnung auf dem DSP, wo der Algorithmus möglichst recheneffizient sein muss.

Codebuch

Codebücher wurden bereits in der Geschichte in vielen Bereichen verwendet, wobei beispielsweise Daten mittels eines Codebuchs verschlüsselt übertragen wurden. Dabei kann

das Codebuch eine Datenbank von Wortzuweisungen sein, so dass beispielsweise *LAF* der Bedeutung *Werden morgen ankommen* zugeordnet ist. Somit kann mithilfe des Codebuches eine Nachricht verschlüsselt übertragen werden. Diese Methode zur Verschlüsselung von Nachrichten wurde beispielsweise im Ersten Weltkrieg angewandt [Ben15]. Ein weiteres Beispiel für ein Codebuch zur Datenübertragung ist das Morsealphabet [Str05, S. 29 ff], welches die Buchstaben in Morsezeichen beschreibt.

Im Allgemeinen werden Codebücher für zwei Bereiche verwendet, für die Datenübertragung, wie beispielsweise die Verschlüsselung oder der Morse-Code, oder zur Datenkompression [Jun+13, 1-198 ff]. Bei der Datenkompression werden beispielsweise die wichtigsten Informationen aus den Daten gefiltert und diese komprimiert in einem Codebuch gespeichert. Die Codebucheinträge können daraufhin wieder als Datenbank genutzt werden.

Bei der Kommunikationseinheit für Atemvollschutzmasken wird das Codebuch zur Mustererkennung genutzt, wobei das Codebuch die komprimierten Informationen aus dem Mikrofonsignal enthält. Das Codebuch ist dabei in die fünf Klassen Nebensprecher, Pause, Ausatmen, Einatmen und Sprachen unterteilt, damit bei der Mustererkennung unter diesen Zuständen unterschieden werden kann. Die einzelnen Klassen können verschieden viele Codebucheinträge haben, welche jeweils einem Vektor mit den extrahierten Informationen aus dem Mikrofonsignal entsprechen, welches in der Merkmalsextraktion beschrieben wird. Dabei werden die einzelnen Codebucheinträge mit $\mathbf{c}_{\mathrm{b,e}}$ bezeichnet, wobei e der jeweilige Vektor des Codebuchs aus der Klasse b ist. Die Erstellung des Codebuchs benötigt eine große Menge an Mikrofonrohdaten von verschiedenen Sprechern. Diese Mikrofonrohdaten werden dafür in die fünf Klassen klassifiziert, so dass jeder Zeitpunkt der Mikrofonrohdaten eine jeweilige Klasse zu geordnet ist. Mittels dieser Klassifizierung kann das Codebuch erstellt werden.

Erstellung des Codebuchs

Für die Erstellung des Codebuchs ist es wichtig, dass die klassifizierten Mikrofonrohdaten der gleichen Merkmalsextraktion, wie während des Betriebes der Kommunikationseinheit, unterzogen werden. Somit werden die Merkmale aus den Rohdaten gemäß der Verarbeitung aus Abschnitt 4.2.2 extrahiert und gespeichert. Auf Basis dieser extrahierten Merkmale kann das Training des Codebuchs vollzogen werden.

Für die Erstellung des Codebuchs ist die Verteilung der Vektoren normalerweise nicht bekannt, weshalb diese bestimmt werden müssen. Diese Bestimmung erfolgt durch eine Häufigkeitsverteilung, wodurch eine möglichst große Anzahl der Trainingsvektoren vorhanden sein sollte. Für das Training des Codebuchs werden iterative Verfahren wie der *K-means*-Algorithmus [Fin13] oder der Linde-Buzo-Gray-Algorithmus (LBG-Algorithmus) genutzt [Hab99; Spe10]. Bei diesen iterativen Verfahren kann eine gute Näherung gefunden werden.

K-Means-Algorithmus

Mit dem *K-means*-Algorithmus wird aus einer gegebenen Menge von Trainingsvektoren ein Codebuch mit N_{B} Codebucheinträgen erzeugt. Der Algorithmus wird in fünf Phasen beschrieben:

1. Initialisierung: Wählen N_{B} beliebiger Trainingsvektoren als Zentren z_{i} der *Cluster*.

2. **Distanzberechnung:** Berechnen der quadratischen euklidischen Distanz d_e zwischen z_i und den Trainingsvektoren.

3. **Klassifizierung:** Zuordnen der Trainingsvektoren zu dem Zentrum z_i mit der kleinsten Distanz d_e.

4. ***Cluster*-Bildung:** Berechnen eines neuen Clusterzentrums z_i durch Mittelwertbildung aller vorhandener Trainingsvektoren im *Cluster* i.

5. **Schwellenüberprüfung:** Berechnen des Mittelwertes der quadratischen euklidischen Distanzen e_m jedes *Clusters*. Daraufhin folgende Überprüfung, ob der Mittelwert von allen N_B Mittelwerten e_m kleiner ist als der Schwellenwert E_m. Falls

$$E_\mathrm{m} \geq \frac{1}{N_\mathrm{B}} \sum_{m=0}^{N_\mathrm{B}-1} e_\mathrm{m} \tag{4.22}$$

   ist, ist das finale Codebuch durch die Zentren z_i gefunden. Andernfalls werden die Phasen 2 bis 5 solange wiederholt, bis die Gl. 4.22 zu trifft.

Ein Nachteil des *K-Means*-Algorithmus ist die langsame Konvergenz sowie die Möglichkeit, in einem relativ schlechten lokalen Minimum des jeweiligen *Clusters* zu bleiben, welches aus den zufälligen Trainingsvektoren als Anfangsbedingung resultiert.

Linde-Buzo-Gray-Algorithmus

Der LBG-Algorithmus ist sehr ähnlich zu dem *K-Means*-Algorithmus, wobei der LBG-Algorithmus gegenüber dem K-Means-Algorithmus besser konvergiert [Fin13, S. 59 ff]. Die bessere Konvergenz resultiert aus der Anfangsbedingung, wobei dem Codebuch nur ein Codebuchvektor zugeordnet ist und somit die Größe des Codebuchs $N_\mathrm{B} = 1$ ist. Das Codebuch der gewünschten Größe wird dabei schrittweise erzeugt. Der LBG-Algorithmus besteht aus sieben Phasen:

1. **Initialisierung:** Der initiale Codebuchvektor wird durch das Zentrum aller Trainingsvektoren bestimmt.

2. **Splitting:** Die Anzahl der Codebuchvektoren wird verdoppelt, indem ein betragsmäßig kleiner Vektor von den aktuellen Codebuchvektoren addiert bzw. subtrahiert wird [Hab99, S. 54 f].

3. **Distanzberechnung, 4. Klassifizierung, 5. Cluster-Bildung, 6. Schwellenüberprüfung:** Die Punkte 2 bis 5 des K-Means-Algorithmus werden durchgeführt, bis die Gl. 4.22 zutrifft.

7. **Iteration in Bezug auf die Codebuchgröße:** Die Schritte 2 bis 6 werden solange wiederholt, bis das Codebuch die gewünschte Größe erreicht hat.

Der LBG-Algorithmus unterscheidet sich somit von dem *K-Means*-Algorithmus nur durch die Initialisierung und die Iteration zur Vergrößerung des Codebuchs. Bei geeigneter Wahl der initialen Codebuchvektoren beim K-Means-Algorithmus werden ähnliche Resultate erzielt.

Anwendung des Codebuchs als Mustererkenner

Das Codebuch vergleicht die Einträge des trainierten Codebuchs mit dem aktuellem Merkmalsvektor [Fre+16; Hut96]. Der Codebuchvektor mit der minimalen quadratischen euklidischen Distanz ist am wahrscheinlichsten, so dass die Klasse dieses Codebuchvektors angenommen wird. Mit dieser Entscheidung können die Klassen Geräusch, Pause, Einatmen, Ausatmen und Sprache erkannt werden.

Die Anzahl der Codebuchvektoren der einzelnen Klassen unterscheiden sich, da das Codebuch für die Klasse Sprache eine größere Varianz im Merkmalsraum aufweist [Bro+15]. Bei den übrigen Klassen ist die Varianz nicht so hoch, da beispielsweise beim Atmen nur die Intensität verändert werden kann, allerdings nicht die spektralen Anteile. Somit hat die Klasse Sprache 64 Codebuchvektoren und für die übrigen Klassen, Geräusch, Pause, Ausatmen und Einatmen, werden jeweils 16 Codebuchvektoren verwendet. Diese Größen haben sich bei den Erkennungsraten als sehr gut gezeigt. Je größer die Klassen sind, desto größer ist der Rechenaufwand. Es muss grundsätzlich beachtet werden, dass die Größen der Codebücher so gering wie möglich gewählt werden, während die Erkennungsrate ausreichend hoch sein muss.

Mit dem trainiertem Codebuch kann die Aktivität der Klassen bestimmt werden. Im Folgenden werden die Klassen der Menge K zugeordnet, damit dieses mathematisch beschrieben werden kann. Zu dieser Bestimmung wird die minimale quadratische euklidische Distanz

$$d_{\mathrm{E},b}\Big(\tilde{\boldsymbol{X}}(k), \mathbf{c}_{\mathrm{b,e}}, k\Big) = \min_{e=0\,\ldots\,N_\mathrm{B}-1} \left\{ \sqrt{\sum_{m=0}^{N_\mathrm{M}-1} \Big(\tilde{X}(m,k) - c_{b,e}(m)\Big)^2} \right\}, \text{ für } b \in K \quad (4.23)$$

zwischen dem Merkmalsvektor $\tilde{\boldsymbol{X}}(k)$ und den Codebucheinträgen $\mathbf{c}_{\mathrm{b,e}}$ berechnet [Sch+12, S. 227]. Die fünf berechneten minimalen quadratischen euklidischen Distanzen der Klassen werden miteinander verglichen und das Argument des Minimums wird bestimmt:

$$d_{\min}(k) = \operatorname*{argmin}_{b\in K} \Big\{ d_{\mathrm{E},b}\Big(\tilde{\boldsymbol{X}}(k), \mathbf{c}_{\mathrm{b,e}}, k\Big) \Big\}. \quad (4.24)$$

Anhand der minimalen Distanz $d_{\min}(k)$ kann nun ermittelt werden, welches Element der Menge K aktiv ist. Der Signalflussgraph des Codebuchs für eine Klasse inklusive der Merkmalsextraktion ist beispielhaft in Abb. 4.9 dargestellt.

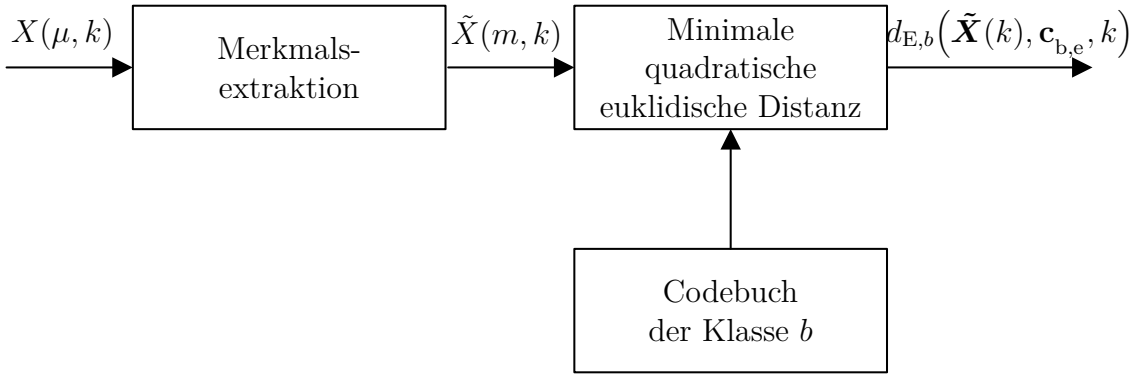


Abbildung 4.9: Signalflussgraph des Codebuchs einer Klasse.

Dieser Signalflussgraph zeigt das spektrale Eingangssignal $X(\mu, k)$ von dem die Merkmale extrahiert werden. Von diesen Merkmalen wird daraufhin mit den Einträgen des Codebuchs der jeweiligen Klasse eine minimale quadratische euklidische Distanz gebildet.

Neuronales Netz

Ein neuronales Netz besteht aus einer netzartigen Verknüpfung von sogenannten Neuronen. Diese sind in verschiedenen Schichten angeordnet. Eine Eingangs- und Ausgangsschicht sind immer vorhanden. Zwischen diesen beiden Schichten gibt es zahlreiche verschiedene Möglichkeiten von verdeckten Schichten [Rey+11]. Die Idee dieser Struktur entstammt im Allgemeinen dem Bestreben die Funktion des Gehirns nachzubilden. Da im Gehirn die Verknüpfungen Neuronen heißen, wird diese Bezeichnung bei den Neuronalen Netzen übernommen [Bis06]. In den Neuronalen Netzen werden die Neuronen, alternativ auch als Knoten bezeichnet, die Bindeglieder zwischen den verschiedenen Schichten. Jedes Neuron kann mehrere Ein- und Ausgänge haben. Die Neuronen sollen die Informationen der Eingänge zu den Ausgängen weiterverteilen. In Abb. 4.10 ist beispielhaft ein neuronales Netz mit Eingangsschicht, verdeckter Schicht und Ausgangsschicht dargestellt.

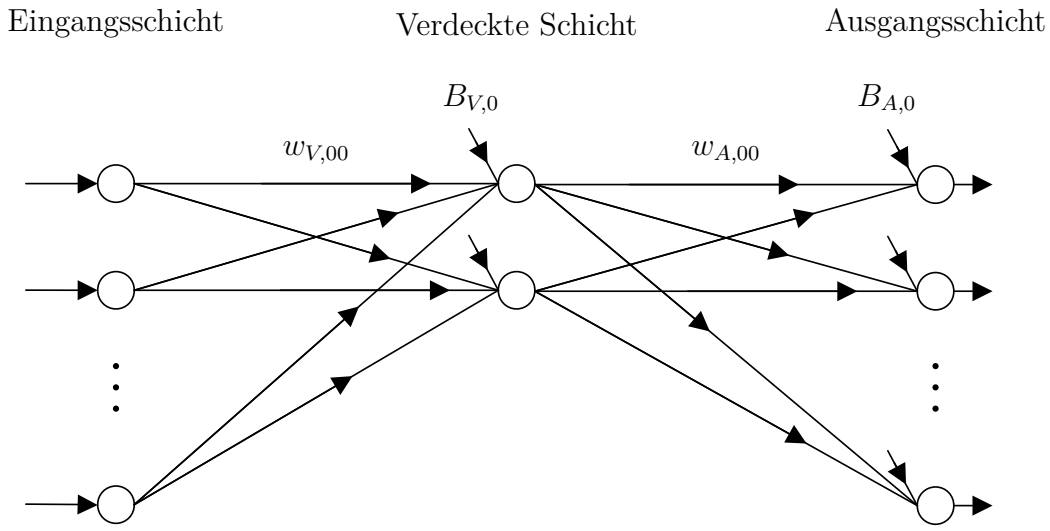


Abbildung 4.10: Beispielhaftes neuronales Netz.

Dabei wird jedes Neuron durch einen Kreis dargestellt. Den Übergängen von Neuron zu Neuron ist jeweils ein Übergangsgewicht zugeordnet. In der Abb. 4.10 sind diese Übergangsgewichte beispielhaft für die obersten Pfade gekennzeichnet [Hof+15]. Die Information eines vorhergegangenen Neurons wird mit einem Übergangsgewicht multipliziert. Daher nimmt der Einfluss des vorhergegangenen Neurons durch ein größer werdendes Übergangsgewicht zu. Bei einem Übergangsgewicht von Null ist kein Einfluss vorhanden. Somit sind die Übergangsgewichte sehr bedeutend, da diese die hauptsächliche Information des Neuronalen Netzes beinhalten. Das Übergangsgewicht w_{ij} ist zwischen dem i-ten sendenden Neuron und dem j-ten empfangenden Neuron. Im Folgenden werden die Formeln allgemeingültig für die Schichten dargestellt, weshalb die Indizes V und A vorerst nicht betrachtet werden. Somit ergibt sich folgender Zusammenhang bei einem empfan-

genden Neuron:

$$y(j) = \sum_{n=0}^{N_j-1} x(n) \cdot w_{nj}, \tag{4.25}$$

wobei x das sendende Neuron, y das empfangende Neuron und N_j die Anzahl der empfangenden Zweige an dem Neuron j ist.

Bei jedem empfangenden Neuron wird zusätzlich zu eingehenden gewichteten sendenden Neuronen ein Bias B hinzuaddiert. Wenn bei den gewichteten sendenden Neuronen kein großer Eingang bei dem empfangenden Neuron besteht, kann das aktuelle Neuron mit einem positiven Bias aktiv und mit einem negativen Bias inaktiv gehalten werden. Somit erweitert sich die Gl. 4.25 um das Bias zu

$$y(j) = B_j + \sum_{n=0}^{N_j-1} x(n) \cdot w_{nj}. \tag{4.26}$$

Aktivierungsfunktion

Bei dem Ausgang eines Neurons der verdeckten oder Ausgangsschicht sind Aktivierungsfunktionen angeordnet. Mit diesen können verschiedene Aktivitätszustände jedes Neurons gesteuert werden. Bei diesen Aktivierungsfunktionen können verschiedene Charakteristiken Verwendung finden:

- die lineare Übertragungsfunktion mit dem Zusammenhang

  $$\mathrm{f}(x) = x, \tag{4.27}$$

  wodurch keine Veränderung resultiert.

- Die begrenzte lineare Übertragungsfunktion wird mit dem Zusammenhang

  $$\mathrm{f}(x) = \begin{cases} \text{oberes Limit}, & \text{wenn } x > \text{oberes Limit}, \\ \text{unteres Limit}, & \text{wenn } x < \text{unteres Limit}, \\ x, & \text{sonst}; \end{cases} \tag{4.28}$$

  beschrieben.

- Eine binäre Übertragungsfunktion mit zwei Zuständen, so dass beispielsweise die Zustände 1 und 0 angenommen werden können:

  $$\mathrm{f}(x) = \begin{cases} \text{oberer Zustand}, & \text{wenn } x > \text{Grenzwert}, \\ \text{unterer Zustand}, & \text{sonst}. \end{cases} \tag{4.29}$$

- Eine sigmoide Übertragungsfunktion mit dem Zusammenhang

  $$\mathrm{f}(x) = \frac{2}{1 + e^{(-2 \cdot x)}} - 1, \tag{4.30}$$

  so dass sich eine Übertragungsfunktion wie in der Abb. 4.11 bildet.

Die sigmoide Übertragungsfunktion wird mit am häufigsten verwendet, durch die Tangens-Hyperbolicus-Funktion ist eine Begrenzung mit einem Minimum und einem Maximum vorhanden [Wal15; Mir+95]. Danach steigt die Kurve langsam an und ist im mittleren Abschnitt näherungsweise linear. Durch diese Charakteristik ist eine bessere Differenzierbarkeit möglich, im Vergleich zu einer binären Übertragungsfunktion. Verglichen mit einer begrenzten linearen Übertragungsfunktion besteht der Unterschied in dem Übergang zwischen der Begrenzung und dem linearen Bereich, bei welchem in der sigmoiden Übertragungsfunktion etwas besser differenziert werden kann. Dabei kann die Bias-Gewichtung wichtig für die Überschreitung oder Unterschreitung einer Grenze der Übertragungsfunktion sein.

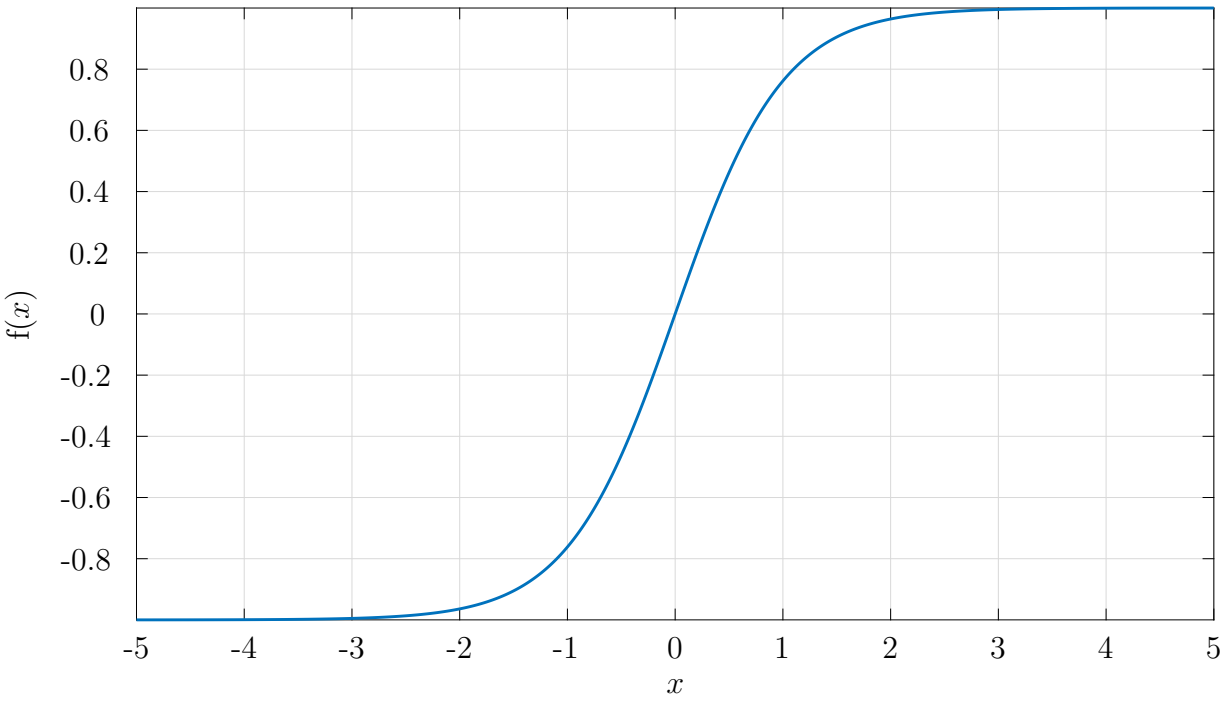


Abbildung 4.11: Sigmoid Übertragungsfunktion.

Verschiedene Neuronale Netztypen

Es bestehen verschiedene Möglichkeiten, ein neuronales Netz zu entwerfen, dabei kann

- das Netz nur Zweige zwischen den Neuronen in der zeitlichen Abfolge besitzen (siehe Abb. 4.10). Diese werden *feed forward*-Netze genannt[Fel+96, S. 153 ff]
- oder das Netz kann Rückkopplungszweige beinhalten, welche von Neuronen zu Neuronen der selben oder einer vergangenen Schicht führen. Diese Netze werden Rekurrente Netze genannt [Rey+11][Abl03, S. 54 ff].

Somit sind die *feed forward*-Netze die Basis der Rekurrenten Netze. Bei den Rekurrenten Netzen wird bei den Rückkopplungen zwischen drei Kategorien unterschieden (siehe Abb. 4.12):

- Der direkte Rückkopplungspfad, welcher den Ausgang eines Neurons mit dessen Eingang verbindet.

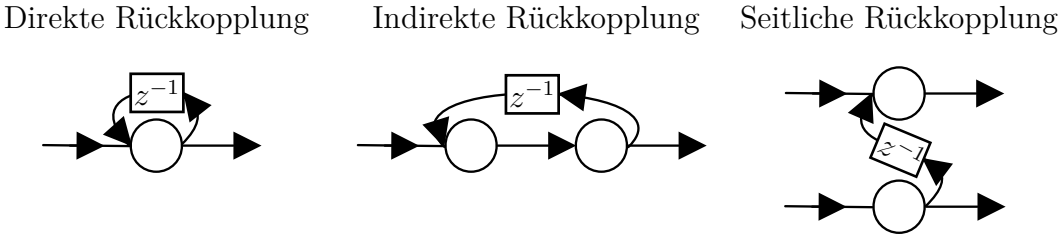


Abbildung 4.12: Verschiedene Rückkopplungszweige (verändert nach [Rey+11]).

- Der indirekte Rückkopplungspfad, welcher den Ausgang eines Neurons mit dem Eingang eines vergangenen Neurons verzweigt.

- Die seitliche Rückkopplung, welcher den Ausgang eines Neurons mit dem Eingang eines im im Parallelzweig liegenden Neurons verknüpft.

Durch die Verwendung von Rückkopplungen kann eine Vorhersage der zukünftigen Zustände gemacht werden. Beispielsweise kann eine Zeitabhängigkeit der Ausgangszustände prognostiziert werden. Somit können bei Zeitabhängigkeiten innerhalb der Mustererkennung die Rückkopplungszweige ein großer Vorteil sein.

Training eines Neuronalen Netzes Für das Training von Neuronalen Netzen sowie Codebücher sind signifikante Trainingsdaten notwendig. Diese werden durch die Merkmalsextraktion vorverarbeitet und daraufhin im Trainingsalgorithmus verarbeitet, so wie es im Kapitel 4.2.4 beschrieben wurde.

Für das Training eines neuronales Netzes sind verschiedene Algorithmen in der Literatur beschrieben, welche beispielsweise

- der Levenberg-Marquardt Algorithmus [Wil+16, S. 12-7 ff] oder

- der *Backpropagation*-Algorithmus [Fel+96, S. 149 ff] sein können.

Für das Training des verwendeten Neuronalen Netzwerkes, wurde der *Backpropagation*-Algorithmus verwendet [Rey+11]. Folglich wird dieser in der folgenden Ausführung näher erläutert.

Der *Backpropagation*-Algorithmus wird beim Training eines neuronales Netzes angewendet, welches aus einer oder mehreren verdeckten Schichten besteht. Es besteht das Problem, dass zwar die Eingangs- und Ausgangsschicht bekannt sind, allerdings nicht die verdeckte Schicht. Um diese Ungewissheit aufzulösen, muss eine den Fehler beinhaltende Gleichung aufgestellt werden.

Für die Lösung dieses Fehlerterms werden beim *Backpropagation*-Algorithmus erst der Eingangsschicht Daten der Trainingsgruppe präsentiert. Damit werden die Ergebnisse der Ausgangsschicht mit den aktuellen Gewichten berechnet. Diese Ergebnisse werden mit den erwarteten Resultaten verglichen und so die begangenen Fehler ermittelt. Falls das ermittelte Ergebnis aufgrund von vorher bestimmten Abbruchkriterien genügt, wird das Training abgebrochen. Wenn die Abbruchkriterien nicht erreicht werden, wird das Training fortgesetzt. Bei der letzten Iteration werden ausgehend von der Ausgangsschicht in

Richtung der Eingangsschicht die Gewichte Schicht für Schicht geändert, so dass der Fehlerterm kleiner wird. Sobald die Modifikation der Gewichte für alle Schichten abgeschlossen ist, werden wiederum am Eingang die Trainingsdaten präsentiert und der Prozess wird solange wiederholt, bis eines der Abbruchkriterien erfüllt ist.

Als optimale Lösung ist ein globales Minimum des Fehlers zu finden. Dieses ist im 2-Dimensionalen eine mit vertretbarem Rechenaufwand lösbare Aufgabe. Durch die Betrachtung des n-dimensionalen Raums, welcher durch ein neuronales Netz mit vielen Neuronen und Schichten entsteht, wird dies deutlich komplexer und dadurch Realisierbarkeit erschwert. Um trotzdem eine sehr gute Lösung für die Auflösung im n-dimensionalen Raum zu erhalten, wird das Gradientenabstiegsverfahren genutzt [Jun13, S. 66 ff][Sch10, S. 94 ff]. Bei diesem Verfahren werden die Gewichte der verschiedenen Schichten modifiziert. Dadurch wird der Fehlerterm iterativ minimiert. Durch die iterative Annäherung ist es nicht notwendig, die Kenntnis der gesamten Hyperebene zu besitzen.

Der Vorteil des Gradientenabstiegsverfahrens ist der geringe Rechenaufwand. Dem gegenüber ist dadurch allerdings nur eine lokale Umgebung bekannt und es wird zumeist nur ein lokales Minimum erreicht.

Beim gewählten Gradientenabstiegsverfahren werden bei der Initialisierung die Gewichte zufällig gewählt. Daraufhin wird für diese Kombination der Gradient bestimmt. Auf Basis dieses Ergebnisses wird versucht, den Fehler zu verringern, welches durch eine Funktion des Gradienten über dem Gewicht erzielt wird. Die Iteration geschieht in einer definierten Schrittweite, wodurch sich die neue Gewichtskombination ergibt. Dieses wird so lange wiederholt, bis ein Abbruchkriterium erreicht wird. Die Abbruchkriterien sind beispielsweise

- eine maximale Anzahl der erlaubten Iterationen,
- eine erlaubte aufeinander folgende Anzahl an Iteration, bei welchen das Fehlerkriterium größer wird und
- der Gradient gleich null ist, welches einem lokalem Minimum entspricht.

Das Gradientenabstiegsverfahren hat durch die iterative Verarbeitung Schwächen, welche beispielsweise folgende Ereignisse sein können:

- Die Kenntnis, ob ein lokales oder ein globales Minimum gefunden wird und wie gut das gefundene lokale Minimum im Vergleich zum globalen ist, ist nicht gegeben. Abhilfe kann hier die Änderung der Startgewichte schaffen, so dass ein optimiertes Minimum gefunden werden kann, welches idealerweise ein globales Minimum ist.
- Eine kleine Steigung des Fehlers zum Gewicht kann dazu führen, dass ein folgendes lokales Minimum nicht erreicht wird, da der Gradient zu klein wird. Abhilfe kann hier eine Erhöhung der Schrittweite sein, so dass die flachen Stellen in der Funktion schneller durchschritten werden und somit das Abbruchkriterium nicht ausgelöst wird.
- Das Verlassen verwendbarer Minima, falls deren Ausprägungen sehr schmal sind und dadurch die Abbruchkriterien nicht auslösen. Im Gegensatz zu den flachen Stellen in der Fehlerfunktion sollte hier die Lernrate reduziert werden, so dass das Minimum gefunden werden kann. Mit der geringen Schrittweite werden bessere Minima

gefunden, allerdings müssen die Abbruchkriterien angepasst werden. Damit kann sich die Trainingszeit deutlich erhöhen.

Mit dieser Trainingsmethode wird das im folgenden beschriebene verwendete neuronale Netz trainiert.

Verwendetes neuronales Netz

Bei diesem Netz sind die Übergangspfade mit einer Gewichtung w_{ij} versehen, welches dem Übergang von den Knoten i zu j entspricht. Dieses *feed forward* Netz besteht aus einer Eingangsschicht, einer verdeckten Schicht und einer Ausgangsschicht. In der Eingangsschicht werden die einzelnen Elemente des Merkmalsvektors $\tilde{X}(m,k)$ mit

$$\tilde{X}(m,k) = 2 \cdot \frac{X_{\mathrm{M}}(m,k) - X_{\mathrm{M,min}}(m)}{X_{\mathrm{M,max}}(m) - X_{\mathrm{M,min}}(m)} - 1 \tag{4.31}$$

auf den Bereich $-1 < \tilde{X}(m,k) < 1$ normiert, wobei m der Merkmalsindex ist. Dabei sind in $X_{\mathrm{M,min}}(m)$ die minimalen und in $X_{\mathrm{M,min}}(m)$ die maximalen Werte des zugehörigen Merkmals der Trainingsdaten, welche zum Training des neuronalen Netzes benutzt werden, hinterlegt. Die Berechnung der Übergänge von der Eingangsschicht zur Ausgangsschicht über die verdeckte Schicht wird gemäß der Gleichung 4.26 durchgeführt. Das Bias der verdeckten Schicht ist B_{Vi} und B_{Aj} für die Ausgangsschicht. Die Übergangsfunktionen sind lineare begrenzte Übertragungsfunktionen mit folgendem Verhalten:

$$\mathrm{f}(x) = \begin{cases} 1, & \text{wenn } x > 1, \\ -1, & \text{wenn } x < -1, \\ x, & \text{sonst.} \end{cases} \tag{4.32}$$

Diese werden genutzt, um die Komplexität bei der Berechnung im Vergleich zu Sigmoid-Funktionen zu reduzieren. Dadurch ergibt sich der Zusammenhang der Übertragung von der Eingangsschicht zur verdeckten Schicht inklusive der linearen begrenzten Übertragungsfunktion gemäß Gl. (4.28) zu

$$X_{\mathrm{V}}(i,k) = \mathrm{f}\left(B_{\mathrm{V}i} + \sum_{n=0}^{N_{\mathrm{M}}-1} \tilde{X}(n,k) \cdot w_{\mathrm{V}in} \right), \text{ für } 0 \leq i < N_{\mathrm{V}}, \tag{4.33}$$

wobei N_{M} die Anzahl der Merkmale, w_{Vin} die Gewichte von der Eingangsschicht zur verdeckten Schicht und $X_V(i,k)$ den anliegenden Wert an dem Knoten i beschreibt. Der Ergebnisvektor der Ausgangsschicht wird mit $\boldsymbol{X}_{\mathrm{A}}(k) = [X_{\mathrm{A}}(0,k),\ ...,\ X_{\mathrm{A}}(N_{\mathrm{A}}-1,k)]^{\mathrm{T}}$ beschrieben und dessen Berechnung ergibt sich aus:

$$X_{\mathrm{A}}(j,k) = \mathrm{f}\left(B_{\mathrm{A}j} + \sum_{i=0}^{N_{\mathrm{V}}-1} X_V(i,k) \cdot w_{\mathrm{A}ji} \right), \text{ für } 0 \leq j < N_{\mathrm{A}}. \tag{4.34}$$

Dabei stellt N_{V} die Anzahl der Knoten der verdeckten Schicht, w_{Aji} das Gewicht von der verdeckten Schicht zur Ausgangsschicht und $X_A(j,k)$ das Ergebnis an dem Knoten j dar. Der Ergebnisvektor der Ausgangsschicht hat $N_{\mathrm{A}} = 5$ Elemente, was der Anzahl der zu erkennenden Klassen entspricht. Dabei wird erneut zwischen Geräusch, Pause, Ausatmen,

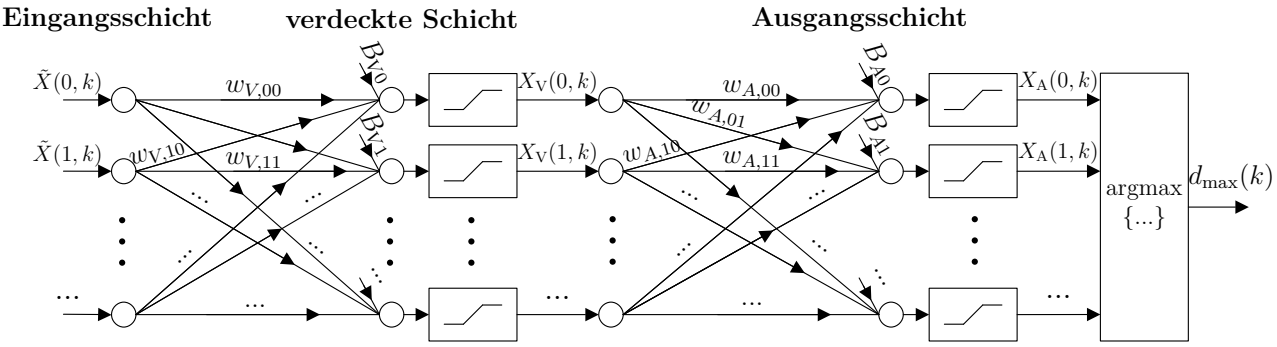


Abbildung 4.13: Verwendetes neuronales Netz.

Einatmen und Sprache unterschieden. Diese werden in der Menge M zusammengefasst, so dass $M = \{\text{Geräusch, Pause, Ausatmen, Einatmen, Sprache}\}$ ist. Zur Detektion, welcher Pfad am wahrscheinlichsten ist, wird der Index des maximalen Eintrags des Vektors $\boldsymbol{X}_{\mathrm{A}}(k)$ bestimmt:

$$d_{\max}(k) = \underset{p \in M}{\operatorname{argmax}} \Big\{ \mathbf{X}_A(p,k) \Big\}. \tag{4.35}$$

Mit dem Index des maximalen Elements $d_{\max}(k)$ kann nun durch das dabei angegebene p ermittelt werden, welches Element der Menge M aktiv ist. Das hier verwendete *feed forward*-Netz ist in Abb. 4.13 zu sehen.

4.2.5 Performance-Vergleich

Performance Vergleich des Codebuchs gegenüber dem neuronalem Netz

Batteriebetriebene Systeme wie die Kommunikationseinheiten für Atemvollschutzmasken werden mit energiesparenden Prozessoren ausgestattet. Um die Batterielaufzeit zu maximieren, sollten alle Algorithmen so effizient wie möglich gestaltet werden. Aus diesem Grund werden die verschiedenen Mustererkennungsansätze hinsichtlich ihrer Erkennungsraten, benötigten Rechenoperationen und auf den benötigten Speicherbedarf hin verglichen.
Die Erkennungsraten des beschriebenen neuronalen Netzes sind in Tab. 4.1 und die des beschriebenen Codebuches in Tab. 4.2 dargestellt, wobei die Klassen Geräusch, Pause, Ausatmen, Einatmen und Sprache durch G, P, A, E und S abgekürzt werden.

		Detection						
		G	P	A	E	S	Erkennung	
Ground Truth	G	53%	23%	14%	2%	8%	53%	47%
	P	14%	70%	8%	1%	7%	70%	30%
	A	2%	2%	83%	1%	12%	83%	17%
	E	<1%	<1%	<1%	99%	<1%	99%	1%
	S	2%	<1%	2%	<1%	95%	95%	5%
	Durchschnittliche Erkennungsrate						79%	21%

Tabelle 4.1: Erkennungsmatrix des neuronalen Netzes.

Ground Truth \ Detection	G	P	A	E	S	Erkennung	
G	38%	40%	9%	3%	10%	38%	62%
P	4%	91%	3%	<1%	2%	91%	9%
A	6%	10%	68%	1%	14%	68%	32%
E	2%	<1%	<1%	96%	<1%	96%	4%
S	3%	3%	5%	<1%	88%	88%	12%
Durchschnittliche Erkennungsrate						76%	24%

Tabelle 4.2: Erkennungsmatrix des Codebuches.

In den Tabellen sind links die Eingangsklassen und oben die Zielklassen abgebildet. Für die Zielklassen sind die Erkennungsraten in % für die jeweilige Eingangsklasse abgebildet; rechts daneben ist die Gesamterkennungsrate für die Eingangsklasse zusammengefasst. Die fehlerhaften Detektionen sind in rot und die richtigen in grün dargestellt. In der untersten Zeile ist die durchschnittliche Erkennungsrate für alle Klassen dargestellt. Bei den Erkennungsraten sollte die Verwechslung zwischen Ausatmen und Sprache möglichst klein sein, da beispielsweise Zischlaute ansonsten als Ausatmen klassifiziert werden.
Der Vergleich der Erkennungsraten beider Mustererkenner zeigt, dass die Erkennungsraten mit dem neuronalem Netz für die Klasse Sprache um 7 %, für Einatmen um 3 %, für Ausatmen um 15 % und für Geräusch um 15 % höher sind. Lediglich die Erkennungsrate für die Klasse Pause ist beim Codebuch um 21 % besser, wobei die Fehlerkennung von Pause beim neuronalem Netz größtenteils in der Klasse Geräusch wiederzufinden ist. Somit werden diese Fehlerkennungen von Pause nicht fälschlicherweise als Sprache detektiert. Die Verwechslung von Ausatmen und Sprache ist beim neuronalem Netz geringer und die durchschnittliche Erkennungsrate um 3 % höher. Somit ist das neuronale Netz in den entscheidenden Fällen dem Codebuch vorzuziehen.
Grundsätzlich müssen die Erkennungsrate in Bezug auf die benötigten Rechenoperationen und den Speicherbedarf gesehen werden, welche für das neuronale Netz und für das Codebuch in Tab. 4.3 dargestellt sind. Der Speicherbedarf ist in Bezug auf die Anzahl der benötigten Parameter im 16 Bit-Format für einen 4 ms Rahmen angegeben.

	Additionen	Multiplikationen	Speicherbedarf
neuronales Netz	188	172	209
Codebuch	3328	1792	1536

Tabelle 4.3: Benötigte Ressourcen der Mustererkenner.

Das neuronale Netz ist im Vergleich zum Codebuch sowohl hinsichtlich der benötigten Rechenleistung als auch des Speicherbedarfs sehr kompakt. Das Codebuch benötigt 17-mal so viele Additionen wie das neuronale Netz, 10-mal so viele Multiplikationen und den 7-fachen Speicherplatz.

Im Gesamtvergleich der Leistungsfähigkeit ist das neuronale Netz gegenüber dem Codebuch merklich im Vorteil, da die Erkennungsraten höher sind, allgemein sowie in den entscheidenden Fällen, und da es deutlich weniger Ressourcen benötigt.

Für eine effizientere Verarbeitung des Codebuchs könnte erst die Klassenzugehörigkeit geschätzt werden und nur bei den signifikantesten Klassen ein vollständige Berechnung erfolgen. Da das Codebuch allerdings schon mit Betrachtung des gesamten Codebuchs keine bessere Erkennung erzielen konnte und das neuronale Netz deutlich weniger Operationen benötigt, wurde das Codebuch nicht weiter optimiert. Durch eine Optimierung kann keine bessere Erkennung erzielt werden.

Mit der hier dargestellten Sprachaktivitätserkennung mit dem neuronalen Netz ist nur die Sprache hörbar, alle Störgeräusche sind unterdrückt. Das Ergebnis ist in der Abb. 4.14b zu sehen; zum Vergleich ist das Sprachsignal ohne Sprachaktivitätserkennung in Abb. 4.14a abgebildet. Diese Aufnahmen sind mit einer realen Kommunikationseinheit durchgeführt worden.
Es ist zu sehen, dass nur die Sprachpassagen nicht gedämpft werden. Die Dämpfung der Nichtsprachpassagen werden durch Ein- und Ausschalten des Verstärkers der Lautsprecher geregelt. Dadurch sind die Nichtsprachpassagen maximal gedämpft. Bei der Sprachaktivitätserkennung ist das *Front-End-Speech-Clipping* sehr wichtig, welches in der Einleitung zu Kapitel 4.2 beschrieben ist. Bei dem konkreten Beispiel ist dieses in der Abb. 4.14b für die Erkennung des Störgeräusches in Sekunde 29 wichtig. Wenn das *Front-End-Speech-Clipping* zu lang ist, wird der Anfang der darauffolgenden Sprachpassage abgeschnitten. Dies kann zu einer erschwerten Verständigung führen. Das *Front-End-Speech-Clipping* ist bei diesem Mustererkenner ca. 20 ms lang. Dies befindet sich im hörbaren Bereich, wodurch der Anfang des ersten Wortes abgeschnitten wird. Die restliche Sprachpassage ist davon nicht betroffen. Es ist besser das *Front-End-Speech-Clipping* so zu akzeptieren, da es aufgrund der Kanalallozierung im Teamfunk über die Sprachaktivitätserkennung wichtig ist, dass alle Störgeräusche unterdrückt werden. Die Sprache des Satzes ist ansonsten davon ausgenommen. Bei Fehldetektionen der Sprachaktivitätsdetektion würde immer der Funkkanal alloziert werden und damit die Performance des Funksystems verschlechtert.

Vergleich der Erkennungsrate eines Mustererkenners mit und ohne linearer Diskriminanzanalyse

Eine Steigerung der Erkennungsrate und eine Dimensionsreduktion der Merkmale kann durch eine Lineare Diskriminanzanalyse erzielt werden, welche im Kapitel 4.2.3 beschrieben ist. Durch eine lineare Diskriminanzanalyse steigert sich der Rechenaufwand durch die Matrixmultiplikation aus der Gl. 4.17. Folglich muss eine Abwägung zwischen dem Rechenaufwand und der Verbesserung der Erkennungsrate geschehen. Hierzu wurden zwei lineare Diskriminanzanalysen durchgeführt. Einmal eine lineare Diskriminanzanalyse ohne Dimensionsreduktion und einmal mit einer Dimensionsreduktion von 12 auf acht Merkmalsvektoren. Bei der linearen Diskriminanzanalyse ohne Dimensionsreduktion kann die Erkennungsrate in jeder Klasse um 1 % gesteigert werden. Der Rechenaufwand steigert sich um 144 Multiplikationen und um 132 Additionen. Bei der linearen Diskriminanzanalyse mit der Dimensionsreduktion ist die Erkennungsrate unverändert und der Rechenaufwand ist um 56 Multiplikationen und 8 Additionen gesteigert. Die lineare Diskriminanzanalyse mit der Dimensionsreduktion erhöht den Rechenaufwand und erzielt keine bessere Erkennungsrate. Bei der linearen Diskriminanzanalyse ohne Dimensionsreduktion ist die Erhöhung der Erkennungsrate im Vergleich zum zusätzlichen Rechenaufwand zu gering, da die Erkennungsrate ohne lineare Diskriminanzanalyse bereits zu einem sehr guten Ergebnis der Sprachaktivitätsdetektion führt, welches in der Abb. 4.14b zu sehen ist. Somit

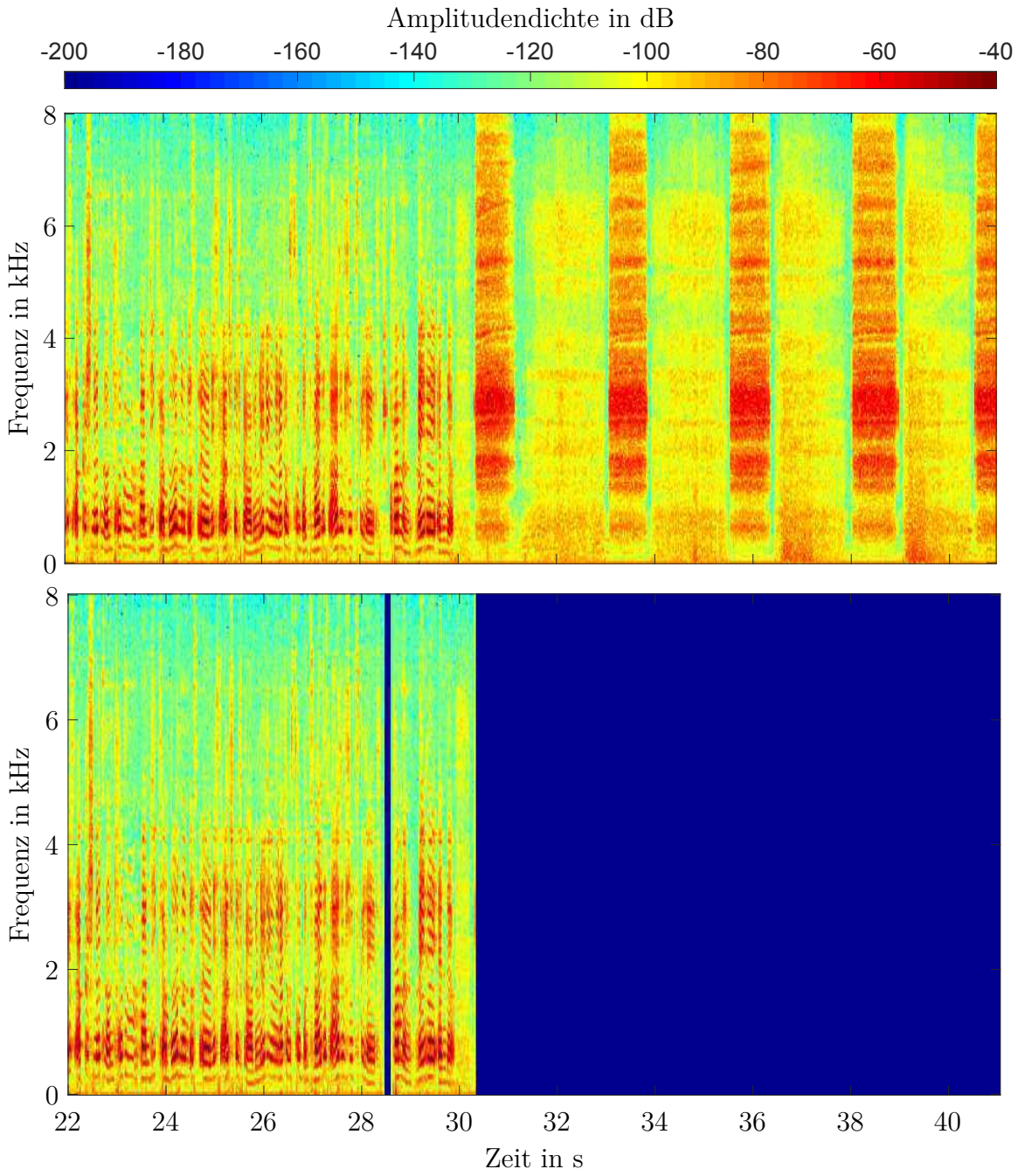


Abbildung 4.14: Spektrogramm von Aufnahmen der Kommunikationseinheit ohne (obere Abb. (a)) und mit (untere Abb. (b)) Sprachaktivitätsdetektion, wodurch nur die Sprachpassagen hörbar sind.

wird die lineare Diskriminanzanalyse bei der Sprachaktivitätsdetektion nicht verwendet.

4.3 Rückkopplungskompensation

Bei der Kommunikationseinheit der Atemschutzvollmaske ist eine Kommunikation über den Teamfunk sehr vorteilhaft. Allerdings ist eine Kommunikation mit dem Sprachverstärker ebenfalls sehr wichtig, da so mit Opfern, Sanitätern oder anderen Einsatzkräften, welche keinen Teamfunk haben, besser kommuniziert werden kann. Bei der Kommunikation mit dem Sprachverstärker gibt es einige Schwierigkeiten, wie die entstehenden Rückkopplungen durch die Nähe von Mikrofon und Lautsprecher und die absolute Lautstärke des Sprachverstärkers. Die absolute Lautstärke wird im Kapitel 5 beschrieben und die entstehenden Rückkopplungen werden in diesem Kapitel behandelt. Die Rückkopplungen bilden sich bei der Kommunikationseinheit der Atemschutzvollmaske besonders stark aus, da das Mikrofon einen sehr kurzen Kopplungsweg vom Lautsprecher zum Mikrofon aufweist. Dieser Weg beträgt auf dem direktem Weg ca. 1 cm und benötigt bei einer Schallgeschwindigkeit von $343,2\,\mathrm{m/s}$ [Gia10] nur $0,3\,\mathrm{ms}$. Durch diesen sehr starken Rückkopplungspfad bilden sich diese besonders stark aus. Zusätzlich befindet sich das Mikrofon auf der Rückseite des Resonanzgehäuses des linken Lautsprechers, so dass der Rückkopplungspfad hier besonders kurz und damit stark ist. Das Gehäuse ist aus Kunststoff und nicht besonders dick, wodurch die Dämpfungseigenschaften eingeschränkt sind. Daher koppelt auf diesem Pfad ebenfalls Schall in das Mikrofon. Hinzu können der linke und rechte Lautsprecher nicht separat ausgesteuert werden, da diese auf einem gemeinsamen elektrischen Audiopfad liegen. Ansonsten hätte der rechte Lautsprecher mit deutlich mehr Leistung betrieben werden können bevor eine Kopplung entsteht, allerdings würde dadurch die Richtcharakteristik der Kommunikationseinheit und die Eigenwahrnehmung des Trägers verändert werden. Die Rückkopplungen werden zusätzlich noch durch die

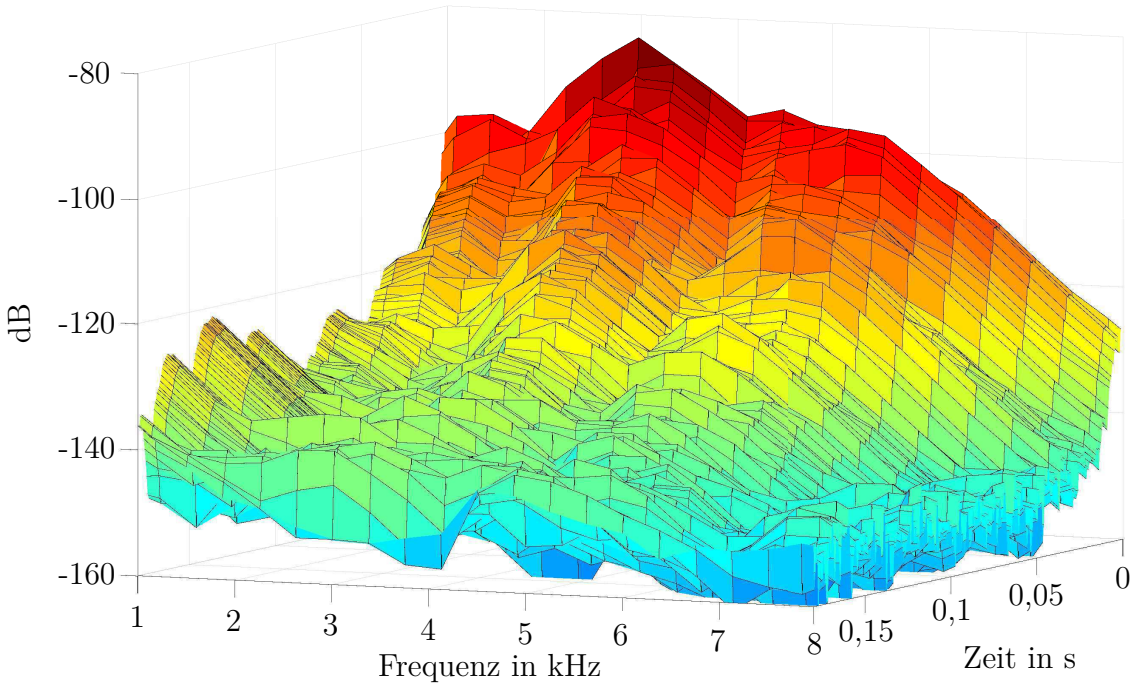


Abbildung 4.15: Breitbandige Anregung der Maske mit Kommunikationseinheit zur Analyse der Ausschwingzeit.

Sprechmembran, welche ein Resonator bei ca. 1 kHz ist, verstärkt, da die Sprechmembran ein sehr langes Ausschwingverhalten hat und somit die Frequenzen immer wieder anregt und sich diese aufschwingen. Das Ausschwingverhalten der Atemschutzvollmaske ist in Abb. 4.15 dargestellt. Hier ist zu sehen, dass die Ausschwingzeit T_{60} frequenzabhängig ist und beispielsweise bei 2 kHz erst nach $T_{60} \approx 100\,\text{ms}$ erreicht wird. Die Ausschwingzeit T_{60} gibt die Zeitspanne an, bis der Schalldruck sich um 60 dB gemindert hat [Wat+08, S. 66]. Somit wirkt sich dieses Verhalten ebenfalls negativ auf die Rückkopplungen aus.
Damit eine ausreichende Lautstärke der Lautsprecher erreicht werden kann, müssen die Rückkopplungen kompensiert werden. Die Kompensation wird mittels adaptiven linearen Kompensationsfilters mit NLMS-Algorithmus erreicht.
Bei der Echokompensation von Freisprecheinrichtung wird auch oft der NLMS-Algorithmus [Ben+13] genutzt, bei der Rückkopplungskompensation besteht allerdings eine starke Korrelation des Lautsprechersignals mit dem Mikrofonsignal, da beide durch den gleichen Sprecher angeregt werden. Diese starke Korrelation führt dazu, dass der Kompensator im Unterschied zu einer Echokompensation nur sehr wenig Verstärkungsgewinn erzielen kann, da die Schätzung des Rückkopplungspfades beeinträchtigt ist. Es kann zusätzlich durch die Korrelation zu Fehlschätzungen kommen und damit zu einem sich aufschwingenden System werden. So wandelt sich der Vorteil des Kompensators zum Nachteil. Aus diesen Gründen muss das Lautsprechersignal dekorreliert werden, so dass ein guter Verstärkungsgewinn durch den Rückkopplungskompensator erzielt werden kann. Die Dekorrelation ist in dem Kapitel 4.3.4 beschrieben. In Abb. 4.16 ist gezeigt, dass dem Kompensator das rückgekoppelte Lautsprechersignal $y(n-d)$ und das Mikrofonsignal $X(\mu, k)$ zur Verfügung stehen. Bei der Implementierung bestand bereits ein Kompensationsalgorithmus, welcher der *Normalized fast block LMS* Algorithmus ist [Hay96, S. 448 ff]. Dieser Blockverarbeitungskompensator war bereits in Festkomma umgesetzt, war in der normalen Anwendung allerdings nicht stabil. Daher werden im folgenden Optimierungsmethoden beschrieben, wodurch der Kompensator einen sehr guten Verstärkungsgewinn erzielen kann. Die vorgenommenen Optimierungen sind eine Schrittweitenkontrolle sowie eine Signaldekorrelation. Um eine bessere Vergleichbarkeit zu haben, werden die verschiedenen Möglichkeiten der Kompensationsalgorithmen verglichen, um aufzuzeigen, wo Verbesserungspotential besteht.

4.3.1 Vergleich der Kompensationsansätze

Bei den adaptiven Algorithmen zur Kompensation sind die beliebtesten Ansätze die Vollbandverarbeitung, die Blockverarbeitung und die Teilbandverarbeitung. Die Vor- und Nachteile sind in der Tabelle 4.4 dargestellt.

	Vollbandverarbeitung	Blockverarbeitung	Teilbandverarbeitung
Auflösung	++ (t) −− (f)	−− (t) ++ (f)	+ (t) + (f)
Rechenaufwand	−−	++	+
Verzögerung	++	−−	−

Tabelle 4.4: Vor- und Nachteile der verschiedenen Kompensationsansätze, wobei in der Auflösungszeile t der Zeitauflösung und f der Frequenzauflösung entspricht (angelehnt an [Hän+04, S.165]).

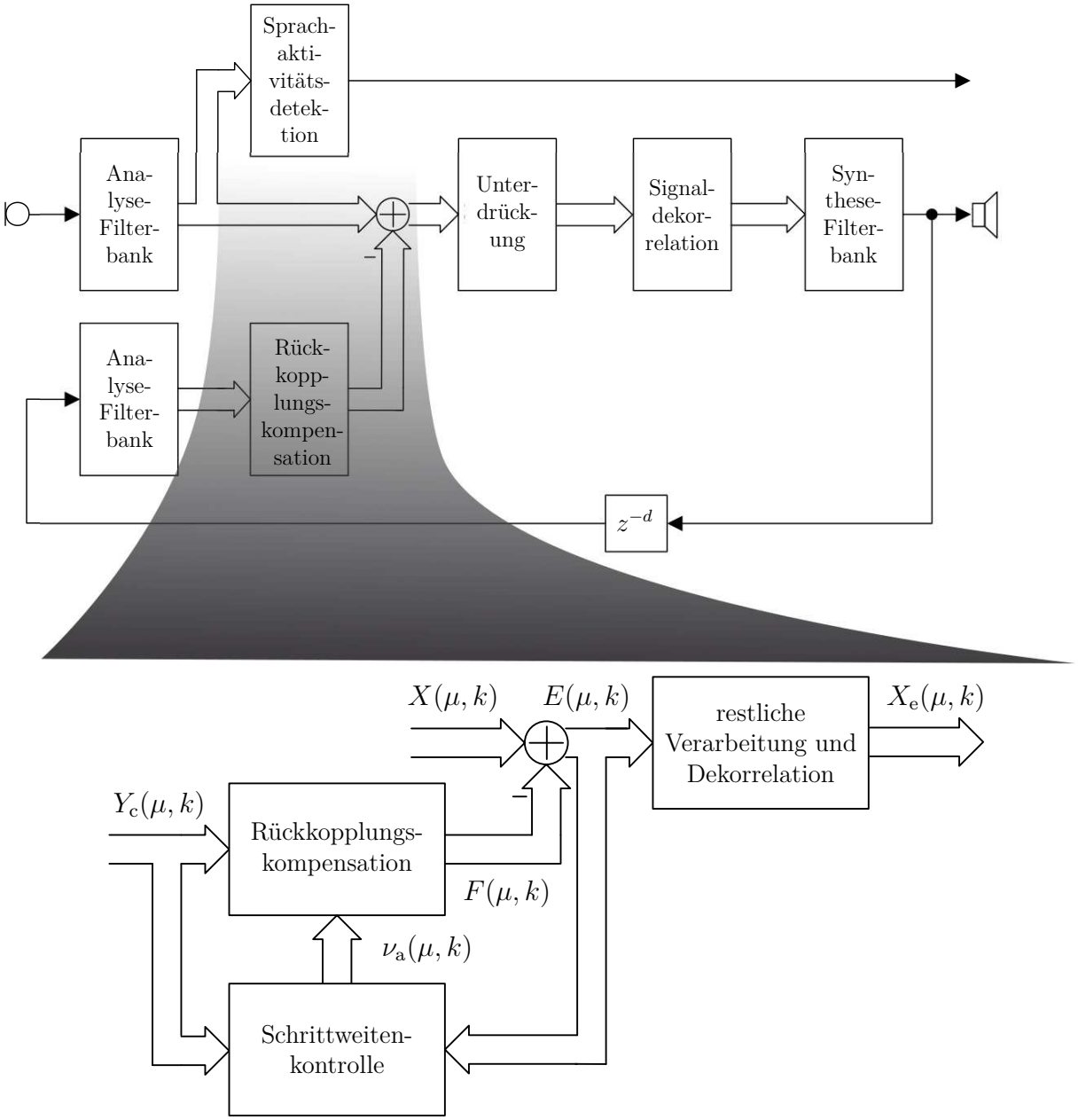


Abbildung 4.16: Struktur der Rückkopplungskompensation mit einer Teilbandverarbeitung (angelehnt an [Hän+04]).

In der Vollbandverarbeitung können die Regelparameter des Algorithmus für beliebige Abtastraten eingestellt werden. Daher ergibt sich eine gute zeitliche Auflösung, wobei allerdings die Frequenzauflösung schlechter ist, da die Regelparameter nicht Frequenzselektiv gewählt werden können. Die Block und Teilbandverarbeitung bieten die Möglichkeit der besseren Frequenzauflösung, aber die Zeitauflösung wird durch Rahmenverarbeitung verringert. Die Vollbandverarbeitung benötigt dabei sehr viele Rechenoperationen, die Blockverarbeitung und Teilbandverarbeitung reduzieren diesen stark, wobei die Blockverarbeitung mit großen Rahmen den geringsten Rechenaufwand benötigt. Damit wird allerdings auch die Verzögerung erhöht. Die Teilbandverarbeitung kann als Kompromiss

zwischen den Ansätzen angesehen werden. In Bezug auf das Zielsystem wäre der Blockverarbeitungsansatz die richtige Wahl, wenn keine anderen Algorithmen im Teilbandbereich genutzt werden. Im Fall der Kommunikationseinheit werden andere Algorithmen im Teilbandbereich genutzt und daher würde nur eine Analysefilterbank zusätzlich benötigt werden und somit wäre der Kompensator mit der Teilbandverarbeitung effizienter. Im vorliegenden System wäre daher die Teilbandverarbeitung am effizientesten. Dadurch, dass die Blockverarbeitung bereits implementiert ist und ungefähr den gleichen Verstärkungsgewinn erzielen kann und die Rechenleistung auf dem bestehenden Prozessor mit der Blockverarbeitung ausreicht, wird im Rahmen dieser Arbeit die Verbesserung der Blockverarbeitung fokussiert. Die Teilbandverarbeitung ist eine Verbesserung, welche genutzt werden kann, wenn die Rechenleistung knapp wird und ist somit als Ausblick zu betrachten.

4.3.2 Blockverarbeitungskompensation

Der verwendete Kompensationsalgorithmus ist der *Normalized Fast Block LMS* Algorithmus, welcher die Eingangssignale des Mikrofons $x(n)$ und das Lautsprechersignal $y(n-\Delta t)$ hat, wobei Δt die Laufzeit vom Lautsprecher bis zum Mikrofon in Takten ist. Das Mikrofonsignal beinhaltet das Nutzsignal des Sprechers sowie das rückgekoppelte Lautsprechersignal als Störanteil. Das Ausgangssignal ist $e(n)$, welches gleichzeitig beim *Normalized Fast Block LMS* das Fehlersignal ist. Als Steuerungssignal wird dem *Normalized Fast Block LMS* die Schrittweite ν_{step} zur Verfügung gestellt, welche im Abschnitt 4.3.3 beschrieben wird. Um die Rückkopplung möglichst gut zu kompensieren, wird versucht die Übertragungsfunktion von Lautsprecher zu Mikrofon zu schätzen. Das Lautsprechersignal wird mit der Übertragungsfunktion multipliziert, so dass dieses Ergebnis von dem Mikrofonsignal subtrahiert wird und als Ergebnis nur das Nutzsignal übrig bleibt. Um dieses Ergebnis zu erhalten, werden zwei Rahmen des Lautsprechersignals der Länge R_{Lsp} aneinandergehängt und davon die FFT gebildet. Es ergibt sich das Signal im Teilbandbereich

$$Y(\mu,k) = \text{FFT}\Big\{[y(n), y(n-1), \cdots, y(n-2R+1)]\Big\}. \tag{4.36}$$

Dieses Lautsprechersignal im Teilbandbereich wird daraufhin durch das geschätzte Leistungsdichtespektrum

$$P_{\text{y}}(\mu,k) = \gamma P_{\text{y}} + (1-\gamma) Y(\mu,k) Y^{\text{H}}(\mu,k) \tag{4.37}$$

normalisiert, wobei γ eine Glättungskonstante ist. Das Filterupdate ergibt sich aus der Berechnung

$$\Phi(\mu,k) = P_{\text{y}}^{-1}(\mu,k) Y^{\text{H}}(\mu,k) E(\mu,k). \tag{4.38}$$

Dabei ergibt sich $E(\mu,k)$ aus dem Fehlersignal $e(n)$ mit dem Zusammenhang

$$E(\mu,k) = \text{IFFT}\Big\{[0_{1\text{ x (B-1)}}, \mathbf{e}^*(n)]\Big\}. \tag{4.39}$$

Daraufhin wird $\Phi(\mu,k)$ in den Zeitbereich transformiert und von dem Blocksignal nur die erste Hälfte der Länge R_{Lsp} erhalten und die zweite Hälfte mit Nullen aufgefüllt.

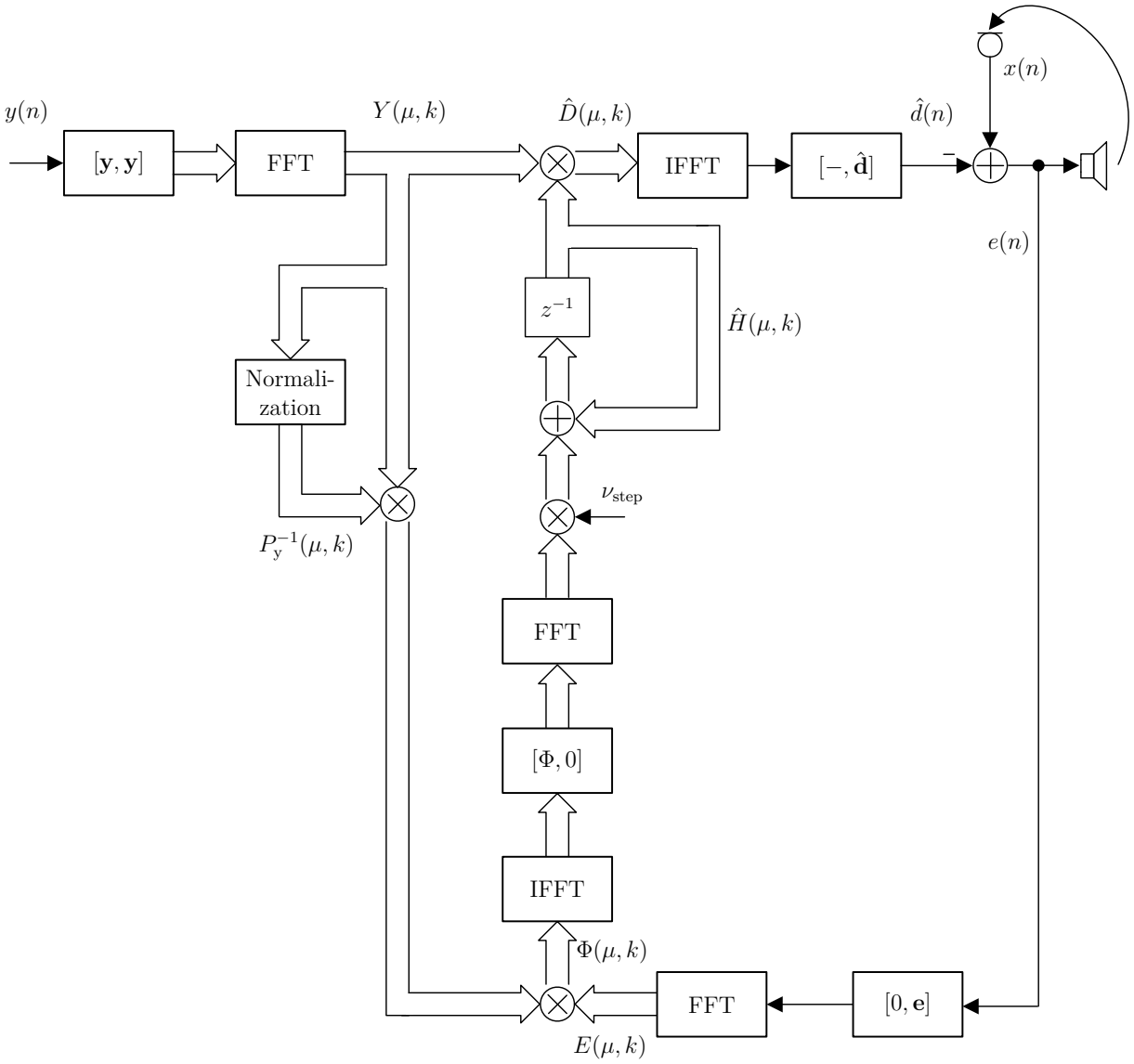


Abbildung 4.17: *Normalized Fast Block LMS* (angelehnt an [Hän+04]).

Dieses Signal wird mit der Schrittweite ν_{step} multipliziert und daraufhin mit der aktuell geschätzten Übertragungsfunktion addiert. Es ergibt sich folgender Zusammenhang:

$$\hat{H}(\mu, k+1) = \hat{H}(\mu, k) + \nu_{\text{step}} \cdot \text{FFT}\Big\{[\phi(n), 0_{1 \text{ x (B-1)}}]\Big\}. \tag{4.40}$$

Mit dieser geschätzten Übertragungsfunktion wird das Lautsprecherspektrum multipliziert, so dass sich das geschätzte Störsignalspektrum $\hat{D}(\mu, k)$ aufgrund der Kopplung ergibt. Dieses Störsignalspektrum wird in den Zeitbereich transferiert und von diesem Signal werden die letzten B Abtastwerte genutzt und ergeben das Signal $\hat{d}(n)$. Dieses Störsignal wird von dem Mikrofonsignal subtrahiert und es ergibt sich das Fehler- bzw. Ausgangssignal

$$\mathbf{e}(n) = \mathbf{x}(n) - \mathbf{d}(n). \tag{4.41}$$

Für die Rahmenlänge R_{Lsp} wird in diesem Algorithmus $R_{\text{Lsp}} = 64$ gewählt, welche einer Zeit von 4 ms entspricht. Durch die FFT-Länge von $2R_{\text{Lsp}}$ wird für die Schätzung der

Übertragungsfunktion ein Zeitfenster von 8 ms genutzt. Dabei muss betrachtet werden, dass die Laufzeit Δt so bestimmt wird, dass die Übertragungsfunktion möglichst gut im 8 ms-Fenster liegt.

4.3.3 Schrittweitenkontrolle

Die Schrittweitenkontrolle ν_{step} und die Glättungskonstante γ sind sehr wichtige Regelgrößen bei dem *Normalized Fast Block LMS* Algorithmus. Mit der Regelgröße γ wird das verzögerte Lautsprechersignal $Y(\mu, k-\Delta t)$ geglättet und damit bestimmt, wie schnell neue Änderungen des Lautsprechersignals für die Adaption der Übertragungsfunktion $\hat{H}(\mu, k)$ betrachtet werden. Mit der Schrittweite ν_{step} wird die Geschwindigkeit der Gesamtadaption der Übertragungsfunktion beeinflusst, wobei das Intervall $0 \leq \nu_{\text{step}} \leq 1$ einzuhalten ist [Hän+04]. Bei der Berechnung des NLMS-Algorithmus kann die optimale Schrittweite bestimmt werden, welche durch

$$\nu_{\text{opt}}(\mu, k) = \frac{\mathrm{E}\left\{\left|E_{\text{u}}(\mu, k)\right|^2\right\}}{\mathrm{E}\left\{\left|E(\mu, k)\right|^2\right\}} \tag{4.42}$$

beschrieben wird (für eine Herleitung siehe [Mad+00]). Das ungestörte Fehlerspektrum ist hierbei durch

$$E_{\text{u}}(\mu, k) = E(\mu, k) - S(\mu, k) - B(\mu, k) \tag{4.43}$$

definiert, wobei $S(\mu, k)$ das Sprachsignalspektrum und $B(\mu, k)$ das Hintergeräuschspektrum ist. Die optimale Schrittweite kann mittels der Kurzzeitleistungsspektren genähert werden

$$\nu_{\text{opt}}(\mu, k) \approx \nu_{\text{step}}(\mu, k) = \frac{P_{e_{\text{u}}}(\mu, k)}{P_e(\mu, k)}, \tag{4.44}$$

dabei wird das Kurzzeitleisunsspektrum durch eine Glättung mittels eines IIR-Filters erster Ordnung geschätzt und mit

$$P_e(\mu, k) = \beta\, P_e(\mu, k-1) + (1-\beta)\left|E(\mu, k)\right|^2 \tag{4.45}$$

und

$$P_y(\mu, k) = \beta\, P_y(\mu, k-1) + (1-\beta)\left|Y(\mu, k-\Delta t)\right|^2 \tag{4.46}$$

berechnet, wobei β die Glättungskonstante und Y das Ausgangssignal nach der Nachverarbeitung vom VA ist. Die Schätzung des ungestörten Kurzzeitleisungsspektrums wird mit dem sogenannten Kopplungsfaktor $c(\mu, k)$ (für genauere Details siehe [Wit17])

$$P_{e_{\text{u}}}(\mu, k) = P_y(\mu, k)\, c(\mu, k) \tag{4.47}$$

berechnet und somit kann die genäherte optimale Schrittweite aus Gleichung 4.44 berechnet werden.

In der Implementierung in der Kommunikationseinheit wird eine feste Schrittweite verwendet, da sich in Versuchen gezeigt hat, dass diese performant genug ist und somit die Komplexität der adaptiven Schrittweite gespart wird. Falls eine höhere stabile Verstärkung notwendig wird, ist die vorher beschriebene adaptive Schrittweite eine Verbesserung, die dann zukünftig implementiert werden kann. Für die für die Kommunikationseinheit gewählte Implementierung der Schrittweite wird zwischen einer Initialisierungs- und einer Laufzeitphase unterschieden. In der Initialisierungsphase wird der 5 Sekunden lange Startton der Kommunikationseinheit abgespielt und in dieser Phase kann schneller adaptiert werden als in der Laufzeitphase. Die Schrittweite in der Initialisierungsphase wird mit ν_{init} angegeben und wird in der Abb. 4.18 mit verschiedenen Konfigurationen dargestellt. Bei dieser Abbildung ist das Fehlersignal gegenüber der Zeit dargestellt. Bei dieser Darstellung ist wichtig, dass der Fehlerterm möglichst klein wird, so dass eine möglichst gute Schätzung der Übertragungsfunktion $\hat{H}(\mu, k)$ geschehen kann. Dabei ist zu sehen, dass ab der Schrittweite $\nu_{\text{init}} = 0,7$ der Adaptionsfehler näherungsweise minimal ist und die Adaption von $\nu_{\text{init}} = 0,5$ nur minimal schlechter ist. Da bei der Startphase auch andere Geräusche $B(\mu, k)$ aktiv sein können, kann das System auch divergieren, so dass die Anpassung sehr schlecht wäre. In diesem Fall wäre eine langsamere Adaption besser und daher wird für die Implementierung $\nu_{\text{init}} = 0,5$ gewählt. Die Laufzeitphase beginnt nach der Initialisierungsphase und der Sprecher in der Maske ist jetzt aktiv. In Abb. 4.19 sind verschiedene Schrittweiten ν_{run} mit vorheriger Initialisierungsphase dargestellt. In dieser Abbildung ist wieder das Fehlersignal über der Zeit dargestellt und ab der 5. Sekunde wird die Schrittweite ν_{run} aktiv. Während der Laufzeit ist eine langsamere Adaption sehr wichtig, da somit einer Divergenz entgegen gewirkt wird. Mit der Schrittweite von $\nu_{\text{run}} = 0,2$ wird eine gute aber nicht zu schnelle Adaption realisiert, so dass bei einer Veränderung der Übertragungsfunktion das System diese noch schnell genug schätzen kann und nicht zu schnell divergiert und somit wird die Schrittweite während der Laufzeit verwendet. Die Auswirkung der Variation der Glättungskonstante γ wird in der Abb. 4.20 dargestellt, wobei die vorher beschriebenen Schrittweiten genutzt werden. Bei dieser Abbildung ist

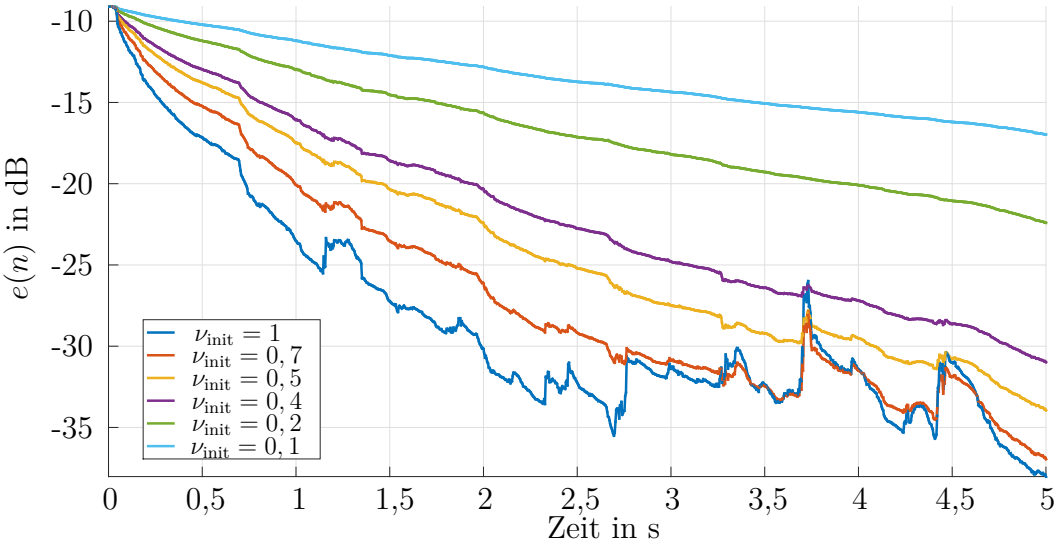


Abbildung 4.18: Darstellung der Schrittweite μ_{init} mit dem Parameter $\gamma = 0,7$

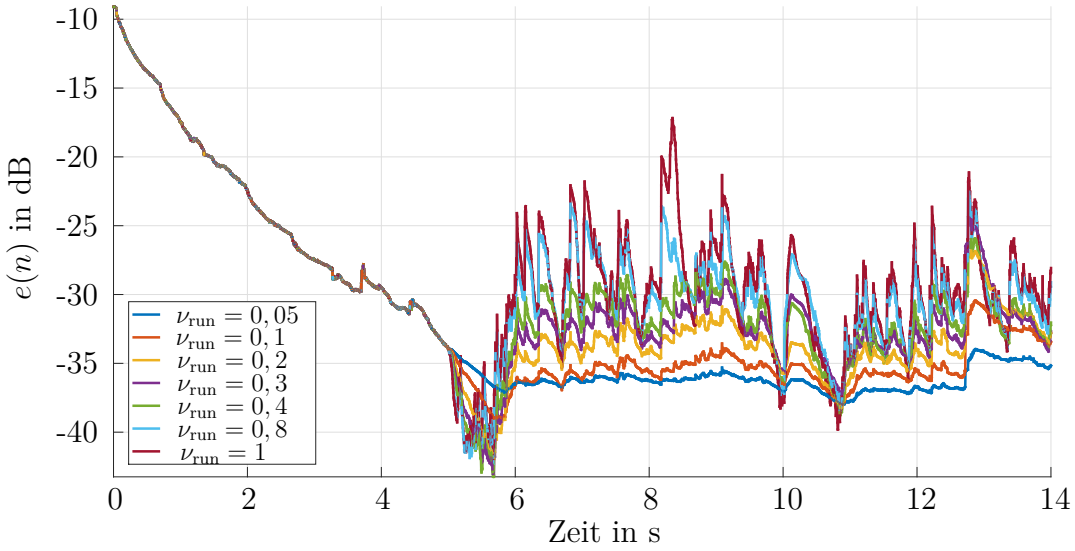


Abbildung 4.19: Darstellung der Schrittweite während der Laufzeit ν_{run} mit den Parametern $\gamma = 0,7$ und $\nu_{\text{init}} = 0,5$. Dabei ist ν_{init} bis 5 s aktiv.

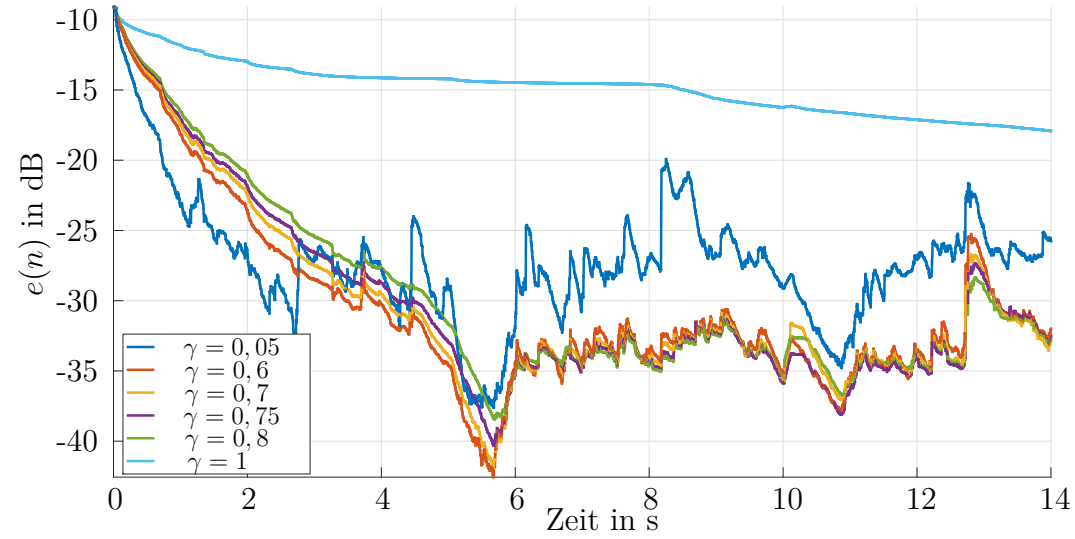


Abbildung 4.20: Darstellung der Glättungskonstante γ mit den Parametern $\nu_{\text{run}} = 0,2$ und $\nu_{\text{init}} = 0,5$. Dabei ist ν_{init} bis 5 s und ν_{run} bis zum Ende aktiv.

wieder das Fehlersignal über der Zeit aufgetragen, um die bestmögliche Konvergenz zu bestimmen. Wenn Gamma zu groß mit $\gamma = 1$ gewählt wird, wird $P_{\text{y}}(\mu, k)$ nach der Initialisierung nicht weiter aktualisiert und damit ist eine gute Konvergenz nicht möglich. Wenn γ zu klein gewählt wird, divergiert das System zu schnell, welches in der Laufzeit zu erkennen ist. Im Bereich $0,6 \leq \gamma \leq 0,8$ konvergiert das System sehr gut. Bei $\gamma = 0,7$ konvergiert das System sehr gut und eine ausreichend schnelle Adaption ist möglich. Somit wurden die Parameter mit $\nu_{\text{init}} = 0,5$, $\nu_{\text{run}} = 0,2$ und $\gamma = 0,7$ gewählt und das System kann die Übertragungsfunktion auch bei schnellen Veränderungen schnell genug schätzen.

Damit bei einer Divergenz das System sich nicht komplett aufschwingt, wird eine einfache Bremse eingebaut, bei welcher die Leistung des Mikrofonsignals $x(n)$ mit der Leistung des Ausgangssignals $e(n)$ verglichen wird. Wenn die Leistung des Ausgangssignals deutlich größer als die Leistung des Mikrofonsignals ist, divergiert das System. Für eine effiziente Umsetzung wird die Bedingung mit der Näherung

$$\sum_{n=0}^{B} |y(n)| \leq \frac{\sum_{n=0}^{B} |e(n)|}{2} \tag{4.48}$$

angenommen. Wenn diese Bedingung erfüllt ist, werden die Stützstellen der Übertragungsfunktion halbiert und für diesen Rahmen ist dann $e(n) = y(n)$ und die Filtergewichte werden erst bei dem nächsten Durchlauf wieder aktualisiert. Somit wird das Aufschwingen verhindert und das System kann danach wieder konvergieren.

4.3.4 Signaldekorrelation

Die Korrelation zwischen dem Lautsprecher- und Mikrofonsignal muss verringert werden, welches durch verschiedene Verfahren der Dekorrelation geschieht. Es wird im folgenden die Dekorrelation mittels eines Frequenzversatzes und die Dekorrelation mittels einer Verzögerung des Lautsprechersignals vorgestellt. Beide Verfahren werden auf das Lautsprechersignal des Sprachverstärkers angewendet. In den folgenden Kapiteln werden die zwei Dekorrelationsverfahren hinsichtlich deren Kohärenz und dessen Wirkung auf den Rückkopplungskompensator untersucht.

Frequenzversatz als Verfahren zur Signaldekorrelation

Verfahren zur Umsetzung eines Frequenzversatzes werden in verschiedenen Publikationen betrachtet [Guo+12; Wit+14]. Aus anderen Themenbereichen ist eine Frequenzverschiebung als eine Einseitenbandmodulation bekannt, wobei das gesamte Signal verschoben wird. Diese Verschiebung kann durch eine Hilberttransformation durchgeführt werden [Joh13; Mil+13], wobei die Verschiebung durch $e^{j\omega t}$ durchgeführt wird.

Bei der Frequenzverschiebung bei einem Sprachsignal wird dadurch dessen Charakteristik verändert und dieses kann sich negativ auf den Höreindruck auswirken. Wird die Sprache betrachtet, kann es zur Verschiebung des Grundtons und dessen Harmonischen führen, wodurch das Signal ab der verschobenen Frequenz nicht mehr harmonisch ist. Daher muss beim Frequenzversatz sehr auf den Höreindruck geachtet werden. Aus diesen Gründen ist es ratsam, den Frequenzversatz in Teilbereiche aufzuteilen [Wit+14]. Bei tiefen Frequenzen dürfen der Grundton und dessen Harmonische nicht so stark verschoben werden, wie beispielsweise bei höheren Frequenzen. Bezogen auf das menschliche Gehör hat dieses bei niedrigen Frequenzen eine höhere Auflösung als bei höheren Frequenzen. Im Folgenden wird das Verfahren der Frequenzverschiebung vorgestellt und auf dessen Implementierungsaspekte eingegangen. In Bezug auf die Kommunikationseinheit der Atemvollschutzmaske wird das Verfahren auf die Wahrnehmung der Frequenzverschiebung durch einen Hörtest evaluiert. Zusätzlich wird die Kohärenz vom Frequenzversatz mit verschiedenen Verschiebungen gezeigt und somit der Einfluss auf den Rückkopplungskompensator.

Mathematische Beschreibung des Frequenzversatzes

Bei dem Frequenzversatz wird die Phase durch einen $e^{j\omega t}$-Term moduliert und somit die Frequenzverschiebung erzielt. Die Berechnung findet im Spektralbereich statt, dabei ist das Eingangsspektrum für den Frequenzversatz das Ausgangsspektrum des Rückkopplungskompensators. Bei diesem werden noch die restlichen Rückkopplungsanteile entfernt und damit ergibt sich das entzerrte Signal $E_{\mathrm{g}}(\mu, k)$. Die Teilbänder sind dabei $\mu = 0 \ldots N/2 + 1$. Das Ausgangsspektrum des Frequenzversatzes ist

$$X_{\mathrm{e}}(\mu + \Delta(\mu), k) = E_{\mathrm{g}}(\mu, k) \cdot F(\mu, k), \tag{4.49}$$

wobei $F(\mu, k)$ der Term der Phasenmodulation und $\Delta(\mu)$ eine Verschiebung in ein anderes Teilband ist. Diese Phasenmodulation wird durch

$$F(\mu, k) = e^{j \cdot 2\pi \, f_{\mathrm{shift}}(\mu) \frac{kR}{f_{\mathrm{S}}}} \tag{4.50}$$

beschrieben, wobei k der Rahmenindex, R der Rahmenversatz, f_{s} die Abtastrate und $f_{\mathrm{shift}}(\mu)$ die Teilband abhängige Verschiebung in Hz ist. Die Verschiebung in ein anderes Teilband $\Delta(\mu)$ findet gemäß

$$\Delta(\mu) = \left\lfloor f_{\mathrm{shift}}(\mu) \, \frac{N_{\mathrm{Sub}}}{f_{\mathrm{s}}} \right\rceil \tag{4.51}$$

statt. Somit wird zum nächstgelegenen Integer-Wert gerundet. Dieses führt zu einem besseren Höreindruck, da somit in das Teilband geschoben wird, welches am nächsten an der gewünschten Frequenz liegt.

Umsetzung des Frequenzversatzes

Bei der Umsetzung des Frequenzversatzes ist es sehr wichtig, dass die Phase korrekt berechnet wird, da ansonsten der Höreindruck beeinträchtigt wird. Der Rechenaufwand des Frequenzversatzes ist bedingt durch den $e^{j\omega t}$ Term sehr hoch, da dieser in einem Prozessor approximiert wird. Diese Approximation ist meist sehr aufwändig, dafür aber genau. Zur Reduzierung des Rechenaufwandes kann der Exponentialterm rekursiv bestimmt werden:

$$F(\mu, k) = F(\mu, k-1) \cdot e^{j \cdot 2\pi \, f_{\mathrm{shift}}(\mu) \frac{R}{f_{\mathrm{S}}}}; \tag{4.52}$$

somit wird der Rahmenindex durch die Multiplikation inkrementiert (vergleiche 4.50). Für die Initialisierung wird der Rahmenindex auf Null gesetzt und somit ist $F(\mu, 0) = 1$ und in jedem Rahmen wird der Phasenmodulationsterm F aktualisiert. Der Exponentialterm kann durch eine Konstante gespeichert werden. Somit ist nur noch eine Multiplikation mit einer Konstanten notwendig. Der Nachteil besteht in der Ungenauigkeit der Konstanten, wodurch sich ein Fehler bei Multiplikation fortpflanzt. Dieser Fehler kann durch eine Korrekturkonstante, welche in festen Zeitabständen addiert wird, ausgeglichen werden. Um den Rechenaufwand weiter zu reduzieren und die Ungenauigkeit auszublenden, kann der Exponentialterm in einem *Lookup-Table* gespeichert werden. Dabei wird die Periodizität ausgenutzt und eine Periode abspeichert. Somit muss der Phasenmodulationsterm F in jeden Rahmen nur noch gelesen werden und keine Multiplikation jedes Teilbandes durchgeführt werden. Der Nachteil besteht hierbei bei dem größeren Speicherbedarf.
In der Implementierung in der Kommunikationseinheit ist die Umsetzung mit dem *Lookup-Table* gewählt worden.

Kohärenzanalyse

Zur Darstellung der Dekorrelation eines Spektrums wird die Kohärenz genutzt, diese wird mit $\gamma_{\mathrm{X,Y}}(f)$ angegeben. Die Kohärenz stellt das Maß für den Grad der linearen Abhängigkeit von zwei Signalen x und y dar. Die Kohärenz hat folgende Berechnung

$$\gamma_{\mathrm{X,Y}}(f) = \frac{|S_{\mathrm{yx}}(f)|^2}{S_{\mathrm{xx}}(f)\, S_{\mathrm{yy}}(f)} \tag{4.53}$$

und ist abhängig von der Frequenz f. Das Kreuzleistungsspektrum ist $S_{\mathrm{yx}}(f)$, die Autoleistungsspektren sind $S_{\mathrm{xx}}(f)$ und $S_{\mathrm{yy}}(f)$, welche jeweils durch die Fouriertransformation der Kreuzkorrelations- bzw. Autokorrelationsfolgen berechnet werden. Die Kohärenz wird im Wertebereich $\gamma_{\mathrm{X,Y}}(f) \in [0, 1]$ angeben, wobei 1 als vollständige und 0 als keine lineare Abhängigkeit interpretiert wird. Je kleiner $\gamma_{\mathrm{X,Y}}(f)$ ist, desto unkorrelierter sind die Signale x und y [Mer13; Gän+96].

Dekorrelationswirkung des Frequenzversatzes

Der Frequenzversatz wird zur Dekorrelation des Signals genutzt, damit mit dem Rückkopplungskompensator der maximal stabile Verstärkungsfaktor ohne Rückkopplungen erhöht werden kann, welcher auch *gain before acoustic feedback instability* bezeichnet wird [Dav+14, S. 236 ff]. Die Dekorrelation des Spektrums wird mit der Kohärenz analysiert.

Die Kohärenz der Frequenzversätze $f_{\mathrm{shift}}(\mu) = 10\,\mathrm{Hz}$, $f_{\mathrm{shift}}(\mu) = 30\,\mathrm{Hz}$ und $f_{\mathrm{shift}}(\mu) = 986\,\mathrm{Hz}$ sind in der Abbildung 4.21 zu sehen. Die Frequenzverschiebung wird hier für die Teilbänder ab 1 kHz bis 8 kHz verschoben und darunter nicht, da das Signal auf dem vorderen Lautsprecher ab 1 kHz übertragen wird und darunter nicht. Bei dieser Berechnung ist der Rahmenversatz $R = 32$ und die FFT-Ordnung $N_{\mathrm{fft}} = 128$ und f_{shift} ist für die bearbeiteten Teilbänder immer gleich.

In der Abbildung 4.21 dekorreliert bereits eine Verschiebung von wenigen Hertz sehr gut, hier beispielhaft mit 10 Hz angegeben. Bei einem Versatz von 30 Hz strebt die Kohärenz bereits gegen Null und bei einem Versatz von 986 Hz ist das Signal vollständig dekorreliert. Ein Nachteil bei dem Versatz von 986 Hz ist, dass bedingt durch den hohen Versatz die Kohärenz erst ab ca. 1,986 kHz sinkt, welches durch das niedrigste verschobene Teilband von ca. 1kHz bedingt ist. Um den Einfluss des Frequenzversatzes auf die Wahrnehmung zu betrachten, wird im Folgenden ein Hörversuch vorgestellt.

Wahrnehmung des Frequenzversatzes in einem Hörversuch

Bei diesen Hörversuch wurde überprüft, welcher Frequenzversatz einen Einfluss auf die Sprachverständlichkeit hat. Mit diesem Ergebniss kann die beste Frequenzverschiebung ausgewählt werden. In diesem Hörversuch wird eine Kombination aus einem MOS- und CMOS-Testverfahren gewählt, wobei bei dem *Mean Opinion Score* (MOS)-Testverfahren das Signal nach der Skala aus der Tabelle 4.5 bewertet wird [Sch08, S. 5]. Diese Bewertung ist in der ITU-T P.800 standardisiert [IT96] und viele standardisierte objektive Maße, wie beispielsweise das Maß PESQ [IT01] oder das Maß *Perceptual objective listening quality assessment* (Polqa) [IT14], haben sich ebenfalls an die MOS-Skala angelehnt. Bei dem *Comparison mean opinion score*-Testverfahren werden zwei Signal dargestellt und diese bezüglich der Qualität miteinander verglichen [Kea00, S. 152 f]. Dieses Verfahren ist

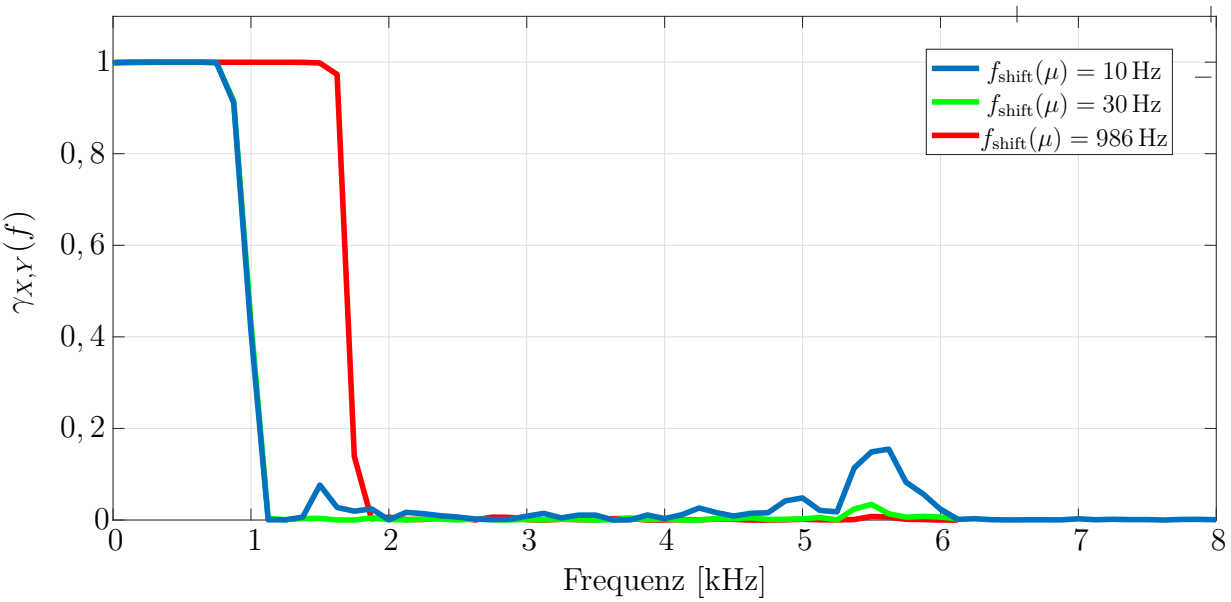


Abbildung 4.21: Kohärenzanalyse der Frequenzverschiebung von $f_{\text{shift}}(\mu) = 10\,\text{Hz}$, $f_{\text{shift}}(\mu) = 30\,\text{Hz}$ und $f_{\text{shift}}(\mu) = 986\,\text{Hz}$

Qualität der Sprache	Bewertungspunkt
ausgezeichnet	5
gut	4
ordentlich	3
mäßig	2
mangelhaft	1

Tabelle 4.5: Beispielhafte MOS-Skala bei der Bewertung der Sprachqualität (nach: [IT96])

ebenfalls in der ITU-T P.800 standardisiert [IT96]. Hierbei sollen Unterschiede der beiden Signale besser dargestellt werden und somit beispielsweise Parametereinstellungen für einen Algorithmus validiert werden. Diese Bewertung ist in der Tabelle 4.6 zu sehen, wobei die Signale mit S1 und S2 gekennzeichnet sind. Bei dieser Bewertung erhält das Signal mit der positiven Bewertung die Punkte positiv und dem anderen Signal werden die Punkte negativ gut geschrieben. In einem typischen CMOS-Test wird immer das Original mit einem veränderten Signal verglichen.

In dem Hörversuch für den Frequenzversatz soll die Sprachverständlichkeit verglichen werden, da es bei der Kommunikationseinheit der Atemvollschutzmaske primär um diese geht. Daher wird in diesem Versuch der MOS- und der CMOS-Test kombiniert. Daraus ergibt sich die Bewertung in der Tabelle 4.7. Hierbei sind die Punkte aus dem MOS-Test gewählt worden und die Bewertungskriterien sind auf die Sprachverständlichkeit ausgelegt. Es wird immer das Original und ein Signal mit einem Frequenzversatz abgespielt und diese sollen zueinander bewertet werden. Die Reihenfolge des Abspielens ist dabei zufällig gewählt. Der Vergleich Original zu verändertem Signal wurde gewählt, da davon ausgegangen werden kann, dass die Sprachverständlichkeit durch den Frequenzversatz nicht gesteigert werden kann, sondern nur beeinträchtigt. Getestet wurde auf der Grundlage

Im Vergleich der Qualität der Sprache erscheint/erscheinen	Bewertungspunkt
S1 viel besser als S2	3
S1 besser als S2	2
S1 etwas besser als S2	1
beide Sequenzen gleich	0
S2 etwas besser als S1	-1
S2 besser als S1	-2
S2 viel besser als S1	-3

Tabelle 4.6: Beispielhafte CMOS-Skala bei der Bewertung der Sprachqualität (nach: [IT96])

Verständlichkeit der Sprache im Vergleich von zwei Sequenzen	Bewertungspunkt
beide Sequenzen erscheinen gleichermaßen verständlich	5
eine der Sequenzen erscheint etwas unverständlicher	4
eine der Sequenzen erscheint unverständlicher	3
eine der Sequenzen erscheint deutlich unverständlicher	2
eine der Sequenzen erscheint viel unverständlicher	1

Tabelle 4.7: MOS-Skala bei der Bewertung der Sprachverständlichkeit im Vergleich von Originalsignal und frequenzverschobenen Signal

von vier Sprechern, jeweils zwei weibliche und männliche Sprecher. Es wurden acht verschiedene Frequenzverschiebungen f_{shift} von 10 Hz, 30 Hz, 50 Hz, 123 Hz, 246 Hz, 492 Hz, 984 Hz, 2958 Hz getestet.

Dargestellt werden die erhobenen Daten des subjektiven Hörversuchs mit einem Boxplot [Cle08, S. 55 ff]. Dieser gibt Information über:

- den Oberen Whisker (90 % der Daten sind kleiner oder gleich dieser Grenze);
- das Obere Quartil (75 % der Daten sind kleiner oder gleich dieser Grenze);
- den Mittelwert (durchschnittlich erreichter Datenwert);
- den Median (50 % der Daten sind kleiner oder gleich dieser Grenze);
- das Untere Quartil (25 % der Daten sind kleiner oder gleich dieser Grenze);
- den Unteren Whisker (10 % der Daten sind kleiner oder gleich dieser Grenze).

In Abbildung 4.22 ist beispielhaft ein Boxplot mit den statistischen Informationen dargestellt [Fah+16].

Bei dem Hörversuch wurden 10 Personen, welche regelmäßig Atemvollschutzmasken tragen, für den Test ausgewählt und dabei sollte nur auf die Sprachverständlichkeit geachtet werden. Das Ergebnis des subjektiven Tests ist in der Abb. 4.23 dargestellt. Hier

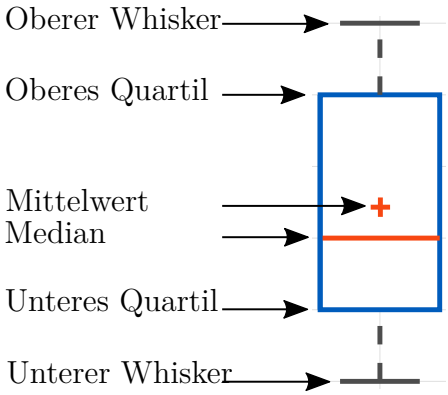
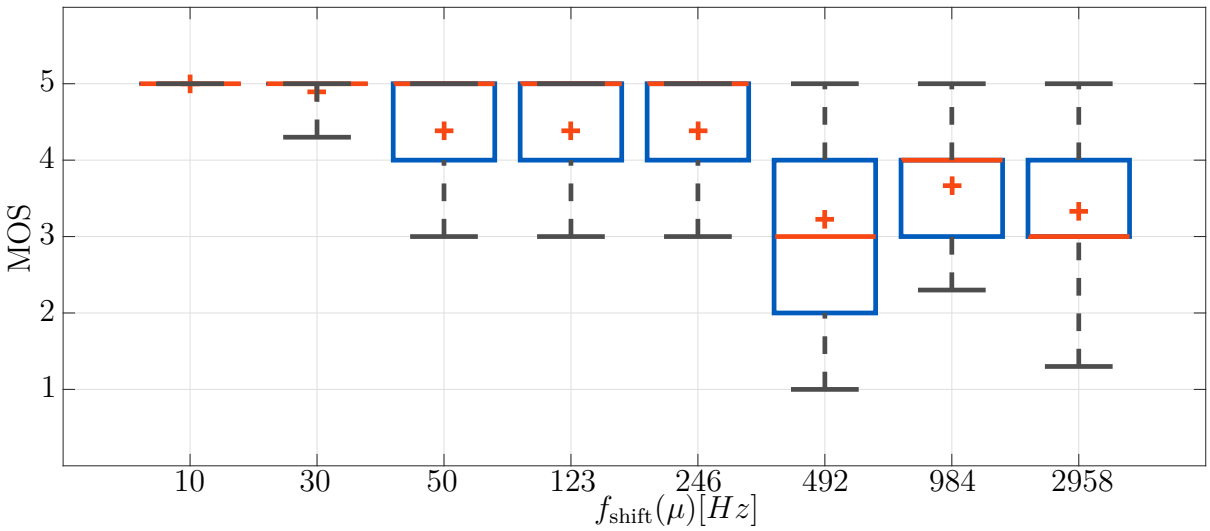


Abbildung 4.22: Darstellung der statistischen Werte in einem Boxplot (nach: [Fah+16])

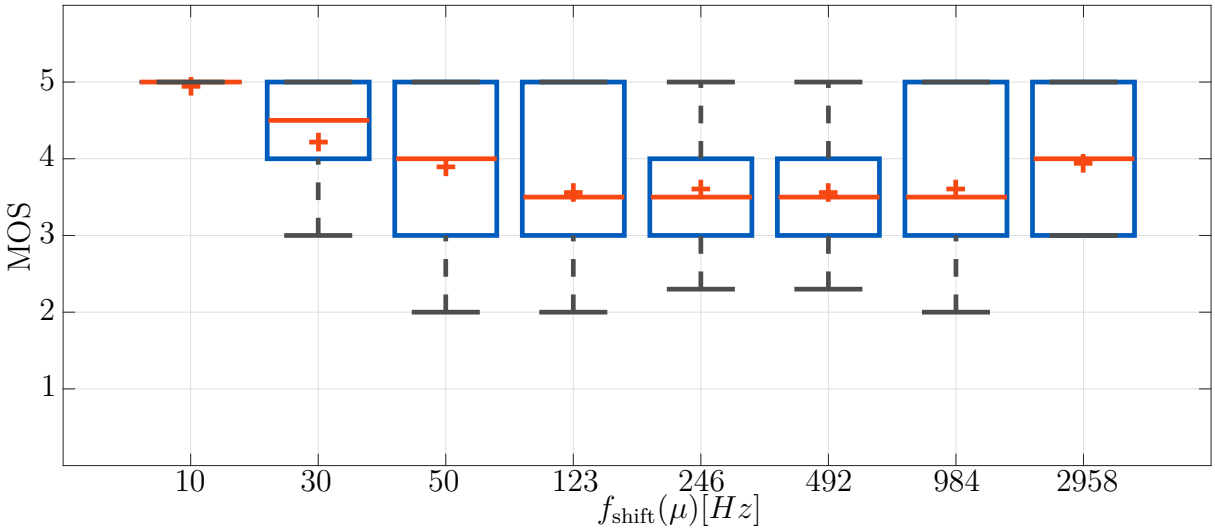
sind die Ergebnisse der männlichen Sprecher in Teil (a) und die weiblichen in Teil (b) dargestellt. Diese Unterscheidung resultiert aus den geschlechterspezifischen Ergebnissen der Frequenzverschiebung, wobei der Unterschied auf die verschiedenen Höhen der Grundfrequenz und der harmonischen Frequenzen zurückzuführen ist. Die erste weibliche Sprecherin hat eine Grundfrequenz f_0 von ca. 180 Hz, die zweite weibliche Sprecherin hat eine Grundfrequenz f_0 von ca. 190 Hz, der erste männliche Sprechern hat eine Grundfrequenz f_0 von ca. 105 Hz und der zweite männliche Sprecher hat eine Grundfrequenz f_0 von ca. 135 Hz. Zusätzlich bilden sich bei den weiblichen Sprecher die Harmonischen der Grundfrequenz bis zu einer höheren Ordnung aus. Daraus resultiert, dass die weiblichen Sprecher durch einen Frequenzversatz, welcher ab 1 kHz angewendet wird, einen größeren Einfluss auf die harmonischen Frequenzen haben. Bei den männlichen Sprechern bilden sich in dem Bereich nur noch wenige Harmonische der Grundfrequenz aus.
Bei einem Frequenzversatz von 10 Hz ist kein Unterschied zu dem Original erkennbar. Bei einem Frequenzversatz von 30 Hz ist bei den männlichen Sprecher ein sehr kleiner Einfluss zu sehen und bei den weiblichen Sprechern ist dieser etwas stärker. Mit einer Erhöhung des Frequenzversatzes über 30 Hz sinkt die Sprachverständlichkeit bei beiden Geschlechtern ab.
Für die Umsetzung ist der Frequenzversatz von 30 Hz auf dem Zielsystem implementiert worden, da die meisten Träger Männer sind und somit eine sehr gute Dekorrelation gewährleistet ist.

4.3.5 Verzögerung

Eine Alternative zu dem Frequenzversatz ist die Dekorralation mittels einer zusätzlichen Verzögerung, durch welche die Signale nicht mehr gleichzeitig am Mikrofon eintreffen und somit dekorreliert sind. Der sofort erkennbare Nachteil ist, dass normalerweise die Laufzeit möglichst gering sein sollte, damit kein Echoeffekt wahrnehmbar ist. Der Echoeffekt tritt laut Literatur ab einer Verzögerung von 30 ms auf [Wei09]. Da die Filterbänke und die DMA-Puffer zusammen eine Laufzeit von ca. 10 ms aufweisen, sollte die zusätzliche Verzögerung maximal 20 ms betragen. Ein weiterer Nachteil gegenüber des Echoeffektes ist, dass die Dekorrelation durch die Verzögerung bei gleichbleibendem Ton nicht eintritt. Bezogen auf Sprache ist die Kohärenz von Signalen mit einer Gesamtverzögerung von 10 ms, 20 ms und 30 ms in der Abb. 4.24 dargestellt. Hier ist zu sehen, dass eine Ver-



(a) MOS-Testergebnisse der männlichen Sprecher



(b) MOS-Testergebnisse der weiblichen Sprecher

Abbildung 4.23: Darstellung der Testergebnisse im Zusammenhang mit der Frequenzverschiebung; (a) Bewertung der männlichen Sprecher, (b) Bewertung der weiblichen Sprecher bei einem Bewertungsumfang von 10 Testpersonen; bewertet wurde die Sprachverständlichkeit.

zögerung von 10 ms bereits sehr gut dekorreliert und die verzögerten Signale um 20 ms und 30 ms nahezu vollkommen dekorreliert sind. Bezieht man diese Ergebnisse auf einen realistischen Einsatz mit der Kommunikationseinheit funktioniert diese Dekorrelation im Normalfall sehr gut, wenn allerdings eine Veränderung der Umgebung beispielsweise durch eine Hand vor dem Lautsprecher verursacht wird, koppelt das System und die Kopplung ist ein stehender Ton und dieser wird durch die Verzögerung nicht dekorreliert. Daher ist die Verzögerung nur bedingt zur Dekorrelation geeignet. Zudem sind dessen Nachteile zu groß und durch den Frequenzversatz wird das Signal bereits sehr stark dekorreliert, daher wird das Verzögerungsverfahren nicht genutzt.

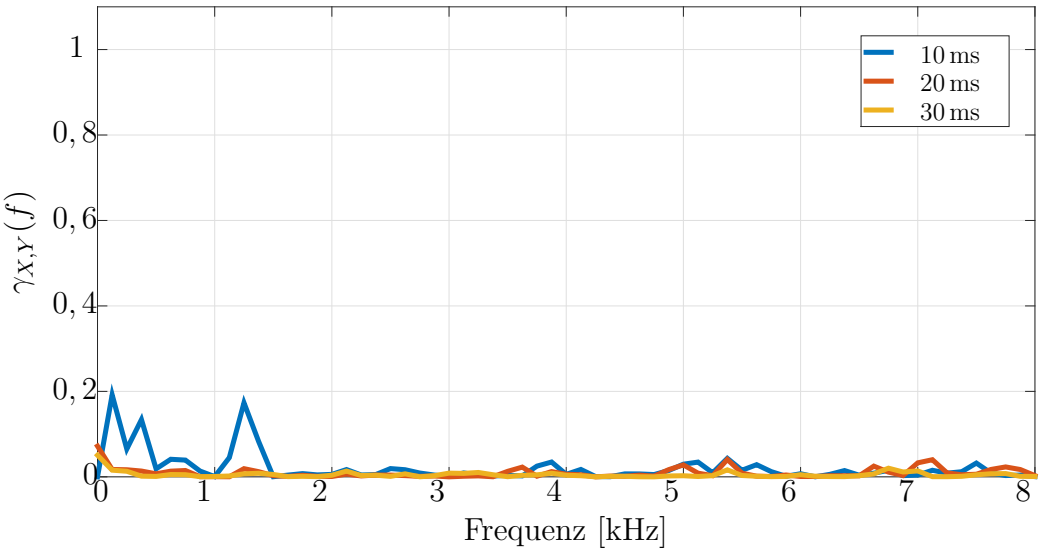


Abbildung 4.24: Kohärenzanalyse bei einer Systemverzögerung von 10 ms, 20 ms und 30ms.

4.4 Geräuschunterdrückung

Die Geräuschunterdrückung ist der Rückkopplungskompensation nachgelagert und minimiert das Hintergrundgeräusch. In Hinsicht auf die Signalverarbeitung ist das Modul in eine Geräuschschätzung und eine Unterdrückung mittels Wiener-Filter aufgeteilt. Die Hintergrundgeräusche werden durch die Geräuschschätzung ermittelt. In der Störsignalunterdrückung wird das Signal der Geräuschschätzung genutzt, damit dieses Störsignal vom Mikrofonsignal mittels eines Wiener-Filters unterdrückt werden kann.

4.4.1 Geräuschschätzung

In der Geräuschschätzung wird das Hintergrundgeräusch geschätzt. Durch die Positionierung des Mikrofons in der Maske sind die Hintergrundgeräusche sehr stark gedämpft, die Geräuschschätzung wird daher nur auf stationäre sehr laute Geräusche ausgelegt wie beispielsweise die Störgeräusche von Schutzbelüftern. Die Geräuschschätzung [Baa12] wird mit dem Mikrofonspektrum nach der Rückkopplungskompensation $E(\mu, k)$ berechnet. Für die Geräuschschätzung wird im ersten Schritt eine genäherte Betragsbildung

$$|E(\mu,k)| = \sqrt{\Re\{E(\mu,k)\}^2 + \Im\{E(\mu,k)\}^2} \tag{4.54}$$

berechnet. Daraufhin wird dieses Signal geglättet, damit die stark zeitvarianten Signale nicht mit betrachtet werden. Die Glättung wird mittels eines IIR-Filters erster Ordnung [Mer13]

$$|\overline{E}(\mu,k)| = \alpha_{\text{sm}} \cdot |E(\mu,k)| + (1-\alpha_{\text{sm}}) \cdot |\overline{E}(\mu,k-1)| \tag{4.55}$$

umgesetzt. Für die Schätzung des Hintergrundgeräuschs wird das Signal $|\overline{E}(\mu,k)|$ mit der Steigungskonstate Δ_{inc} oder der Abfallkonstante Δ_{dec} multipliziert. Für die Schätzung

des Hintergrundgeräuschs wird hierbei Δ_{inc} deutlich kleiner als Δ_{dec} gewählt, somit werden langsam veränderliche Hintergrundgeräusche geschätzt. Der Schätzer folgt durch das langsame Δ_{inc} schnell veränderlichen Signalen nicht. Die durch Δ_{inc} verursachte kleine Steigung wird daraufhin mit Δ_{dec} wieder schnell auf den Pegel des Hintergrundgeräuschs abgesenkt. Die Berechnung für das geschätzte Hintergrundgeräusch $\hat{B}$ erfolgt durch:

$$\hat{B}(\mu,k) = \begin{cases} \Delta_{\text{inc}} \cdot \hat{B}(\mu,k-1), & \text{wenn } |\overline{E}(\mu,k)| > \hat{B}(\mu,k-1), \\ \Delta_{\text{dec}} \cdot \hat{B}(\mu,k-1), & \text{sonst.} \end{cases} \tag{4.56}$$

In anderen Kommunikationssystemen, bei welchen die Hintergrundgeräuschen nicht so stark gedämpft werden, wie bei der Maske, können auch andere Verfahren sinnvoll sein. Beispiele für solche Geräuschschätzungen sind in den Referenzen [Baa12; Ben+09] zu finden.

4.4.2 Wiener-Filter

Das Wiener-Filter ist ein stochastischer Algorithmus zur Rauschunterdrückung [Kov+08], bei welchem die Vorkenntnis der statistischen Eigenschaften der involvierten Signale bis zur zweiten Ordnung erforderlich ist. Bei diesem Filter soll beispielsweise additives Rauschen im Nutzsignal unterdrückt werden. Im Folgenden wird das additive Rauschen mit $b(n)$, das Nutzsignal mit $s(n)$, das Eingangssignal gemäß Abb. 4.1 mit $x(n)$ bezeichnet und der Zusammenhang

$$x(n) = s(n) + b(n) \tag{4.57}$$

verwendet. Mit den Auto- bzw. Kreuzleistungsdichtespektren $S_{\text{bb}}(\Omega), S_{\text{ss}}(\Omega), S_{\text{xx}}(\Omega)$ und $S_{\text{xs}}(\Omega)$ ergibt sich bei der Annahme der Orthogonalität zwischen s und b folgender Zusammenhang (Herleitung siehe [Hän+04]): Mit einer Fourier-Transformation [Ger+97] kann die Autokorrelation und die Kreuzkorrelation in den Frequenzbereich umgewandelt werden:

$$\begin{aligned} H_{\text{opt}}(\Omega) &= \frac{S_{\text{xs}}(\Omega)}{S_{\text{xx}}(\Omega)} \\ &= \frac{S_{\text{ss}}(\Omega)}{S_{\text{xx}}(\Omega)} \\ &= \frac{S_{\text{xx}}(\Omega) - S_{\text{bb}}(\Omega)}{S_{\text{xx}}(\Omega)} \\ &= 1 - \frac{S_{\text{bb}}(\Omega)}{S_{\text{xx}}(\Omega)}. \end{aligned} \tag{4.58}$$

Mit dieser optimalen Übertragungsfunktion wird das Leistungsdichtespektrum des Eingangssignals multipliziert und somit die berechneten Komponenten in H_{opt} unterdrückt. Um eine sehr starke Dämpfung zu unterbinden wird eine Maximaldämpfung H_{min} eingeführt und in die Gleichung (4.58) wie folgt

$$H_{\text{opt}}(\Omega) = \max\left\{H_{\text{min}}, 1 - \frac{S_{\text{bb}}(\Omega)}{S_{\text{xx}}(\Omega)}\right\} \tag{4.59}$$

einbezogen. Zusätzlich sind die Änderungsgeschwindigkeit der Geräuschschätzung und der Sprache unterschiedlich, wobei sich die Geräuschschätzung wesentlich langsamer als die Sprache ändert. Durch diesen Unterschied öffnet und schließt das Filter, woraus sich Töne aus dem restlichen Rauschen ergeben (sogenannte *musical tones*). Um dieses Verhalten zu unterbinden wird ein Überschätzungsfaktor β_{over} in die Gleichung (4.59) folgendermaßen eingebunden:

$$H_{\text{opt}}(\Omega) = \max\left\{H_{\min}, 1 - \frac{\beta_{\text{over}} \cdot S_{\text{bb}}(\Omega)}{S_{\text{xx}}(\Omega)}\right\}. \tag{4.60}$$

In diese Gleichung wird nun das geschätzte Leistungsdichtespektrum des Hintergrundgeräusches

$$\hat{S}_{\text{bb}}(\mu, k) = \hat{B}(\mu, k)^2 \tag{4.61}$$

und das Leistungsdichtesprektrum des Mikrofonsignals

$$\hat{S}_{\text{xx}}(\mu, k) = |E(\mu, k)|^2 \tag{4.62}$$

eingesetzt. In der Implementierung wird $\hat{S}_{\text{xx}}(\mu, k)$ optimiert zu

$$\hat{S}_{\text{xx}}(\mu, k) \approx \left(\left|\Re\Big\{E(\mu, k)\Big\}\right| + \left|\Im\Big\{E(\mu, k)\Big\}\right|\right)^2, \tag{4.63}$$

damit die benötigten Rechenoperationen minimiert werden. Mit $\hat{S}_{\text{bb}}(\mu, k)$ und $\hat{S}_{\text{xx}}(\mu, k)$ ergibt sich

$$H(\mu, k) = \max\left\{H_{\text{min,b}}, 1 - \frac{\beta_b \cdot \hat{S}_{\text{bb}}(\mu, k)}{\hat{S}_{\text{xx}}(\mu, k)}\right\}. \tag{4.64}$$

Mit dem Dämpfungsfaktor $H(\mu, k)$ und dem Mikrofonspektrum $E(\mu, k)$ wird das vom Störgeräusch bereinigte Mikrofonspektrum

$$E_{\text{s}}(\mu, k) = E(\mu, k) \cdot H(\mu, k) \tag{4.65}$$

bestimmt.

In der Kommunikationseinheit ist diese Geräuschunterdrückung mit einer FFT-Ordnung von $N_{\text{fft}} = 128$ und einem Rahmenversatz von $R = 32$ genutzt. Die maximale Dämpfung ist mit $H_{\text{min,b}} = -40\,\text{dB}$ und die Überschätzung ist mit $\beta_b = 4$ eingestellt. Mit diesen Parametern können die Hintergrundgeräusche signifikant gedämpft werden, welches in der Abb. 4.25 dargestellt ist. Dabei ist das Spektrogramm des Sprachsignals ohne Rauschen in der oberen Abbildung, das Signal mit dem überlagerten weißen Rauschen in der Abbildung (b) und das Signal nach der Geräuschunterdrückung in der Abbildung (c) zu sehen. Nach der Geräuschunterdrückung kann das Hintergrundgeräusch um ca. 25 dB reduziert werden, wodurch die Sprachverständlichkeit verbessert wird. Bei einer höheren Unterdrückung werden die ersten Sprachpassagen angegriffen und somit würde dann die Sprachverständlichkeit sinken. Nach der Geräuschunterdrückung wird dieses Signal in die Signaldekorrelation weiter gegeben.

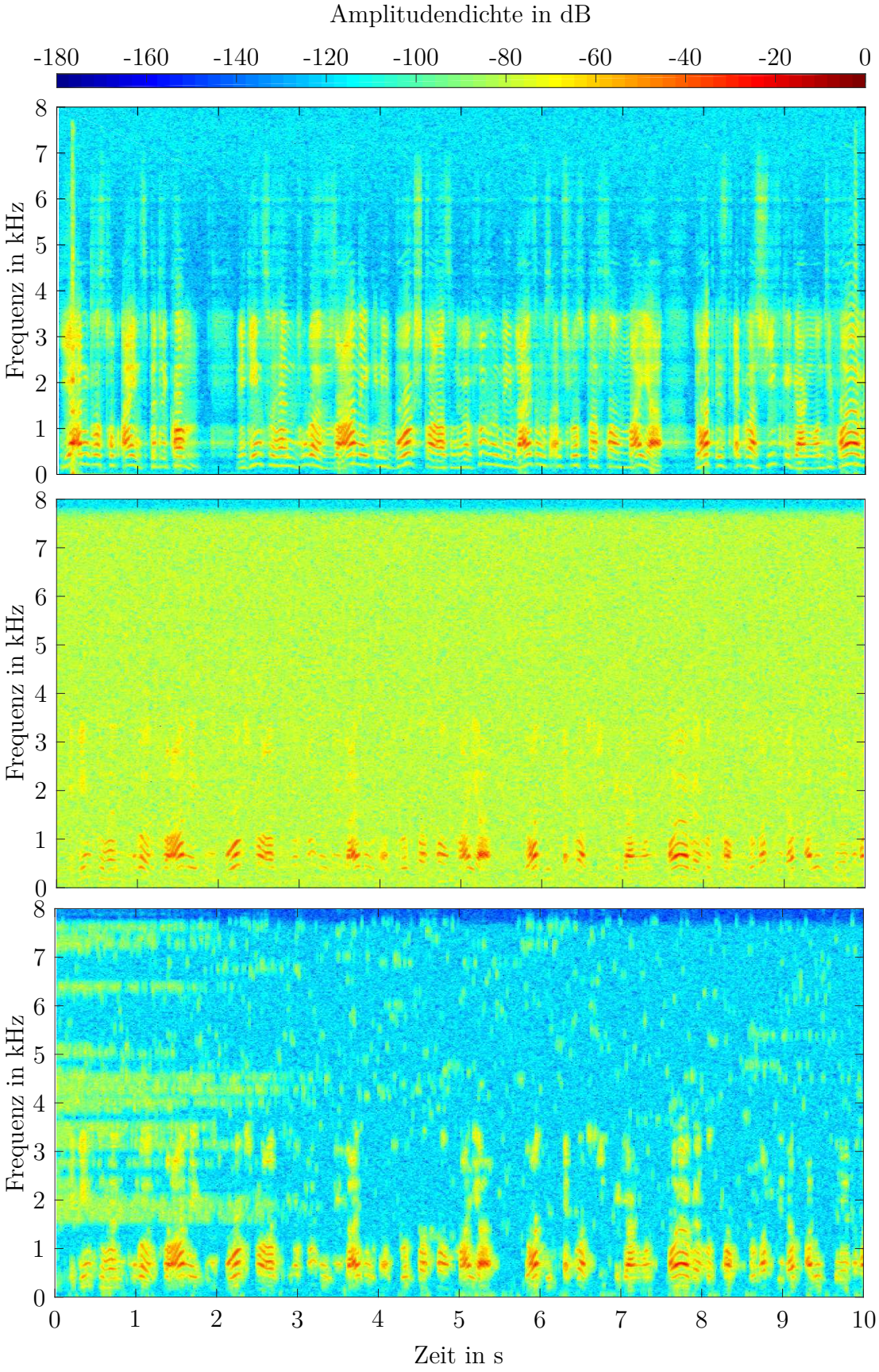


Abbildung 4.25: Spektrogramm des Sprachsignals ohne Rauschen (obere Abb. (a)) und mit weißem Rauschen überlagert (mittlere Abb. (b)) und das Signal nach der Geräuschunterdrückung (untere Abb. (c)).

Kapitel 5

Nachverarbeitung

Die Nachverarbeitung findet nach der Mischung und Verstärkung der Signale statt, dabei beinhaltet die Nachverarbeitung Algorithmen zur Verbesserung der Ausgangssignale des VA's, der Ohrhörer, des Funkgeräteausgangs und des Team-Funk-Ausgangs. Diese Verarbeitung findet ausschließlich im Zeitbereich statt, da somit der Rechenaufwand stark reduziert werden konnte. Wenn in jedem Ausgangskanal die Algorithmen im Frequenzbereich stattfinden, muss für jeden Ausgang eine inverse Fourier-Transformation gerechnet werden, anstatt der aktuellen inversen Fourier-Transformationen in der Mikrofonsignalverbesserung und in der Funksignalverbesserung. Damit beläuft sich nach der Tabelle 2.1 der ersparte Rechenaufwand auf $2 \cdot 7315 = 14630$ Verarbeitungszyklen pro verarbeitetem Rahmen.

Die Nachverarbeitung ist in der Abbildung 5.1 dargestellt, somit ist es bei jedem Ausgangskanal möglich, einen Exciter, einen Equalizer, einen Regelverstärker und einen Hard-Limiter zu benutzen. Für den Funkgeräteausgang wird eine Dynamikanpassung mittels Regelverstärker nicht vorgenommen, da in den Funkgeräten eine automatische Verstärkungsregelung (AGC, *automatic gain control*) integriert ist. Das Eingangssignal der Nachverarbeitung ist $y(n)$, welches der Ausgang der Mischung und Verstärkung ist, dieses Signal wird dem Exciter-Algorithmus präsentiert. Mit dem Exciter-Algorithmus werden mittels nichtlinearer Kennlinien Harmonische der Grundfrequenz erzeugt, welche vorher durch die Dämpfung der Atemvollschutzmaske nicht mehr hörbar sind. Der Equalizer ist ein Biquad-Filter, welcher kaskadiert werden kann, so dass die Frequenzverläufe der Ausgangskanäle so angepasst werden können, wie es erwünscht ist. Mit dem Regelverstärker wird die Dynamik der Signale der Ausgangskanäle angepasst, so dass ein Expander, Durchleiter, Kompressor oder Begrenzer mittels einer Regelverstärker-Kennlinie realisiert werden können. Mit dem Hard-Limiter werden die Signale endgültig begrenzt, so dass beispielsweise ein Schutz der elektronischen Komponenten gewährleistet werden kann. Dieser wird in Verbindung mit dem Expander des Regelverstärkers abgestimmt, so dass die Dynamikbegrenzung mittels des Expanders entsteht und nur in seltenen Fällen der Hard-Limiter eingreifen muss. Somit kann häufiges *Clipping* vermieden werden. Das entstehende Ausgangssignal wird mit $y_a(n)$ gekennzeichnet.

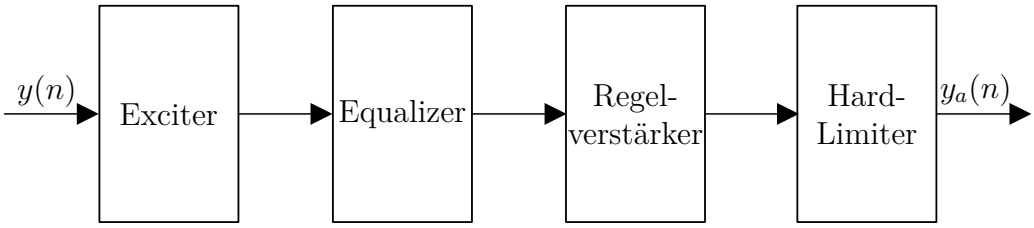


Abbildung 5.1: Signalflussgraph der Nachverarbeitung.

Im Folgenden werden die hier erwähnten Algorithmen vorgestellt.

5.1 Exciter

Der verwendete Exciter wurde 2017 auf der DAGA veröffentlicht [Grö+17]. In der Audiosignalverarbeitung werden nichtlineare Verzerrungen normalerweise als unerwünschter Effekt angesehen, welches beispielsweise bei der Sprachmembran der Atemvollschutzmaske ebenfalls der Fall ist. Im Gegensatz zu dieser Aussage gibt es Anwendungen, bei denen die nichtlineare Verzerrung hinzugefügt wird, um den Klang positiv zu beeinflussen, wobei durch die nichtlineare Verzerrung ganzzahlige Vielfache des Grundtons erzeugt werden sollen. Die ganzzahligen Vielfachen des Grundtons, auch Harmonische genannt, tragen maßgeblich zu der Klangfarbe des Tones bei [Fri14, S. 153 f]. Fehlen Teile der Harmonischen, ist die Klangfarbe verändert und es kann dazu führen, dass beispielsweise ein Ton nicht mehr als dieser wahrgenommen wird. Dieses Phänomen kann am Beispiel eines Röhrenverstärkers gezeigt werden, dieser erzeugt bei hoher Aussteuerung nichtlineare Verzerrungen und daraus resultieren Harmonische. Dieser Effekt wird oftmals als akustisch angenehm empfunden. Dieser Effekt kann ebenfalls mit sogenannten Effektgeräten erzeugt werden, diese gibt es beispielsweise von der Firma Aphex [Ltd01]; der Effekt zur Erzeugung von Harmonischen mittels nichtlinearer Kennlinien wird dort als *Exciter* bezeichnet, woher der Name des Algorithmus dieses Kapitels stammt. Die ersten Effektgeräte wurden in den 70er Jahren entwickelt, um in der Musikproduktion den Gesang und Musikinstrumente mit Effekten zu versehen. Die psychoakustischen Effekte des Aural Exciters der Firma Aphex wurden in der Veröffentlichung von Josef Chalupper untersucht. Er kam zu dem Ergebnis, dass eine Verbesserung der Sprachverständlichkeit durch die Anwendung des Exciters bei Frequenzen über 1 kHz resultieren kann [Cha00]. Ein Beitrag bezüglich der Anwendung des Exciters in der Signalverarbeitung wurde 2016 veröffentlicht [Bul+16], bei nichtlineare Kennlinien genutzt wurden, um die Sprachverständlichkeit von Freisprech-Telefonie zu verbessern.

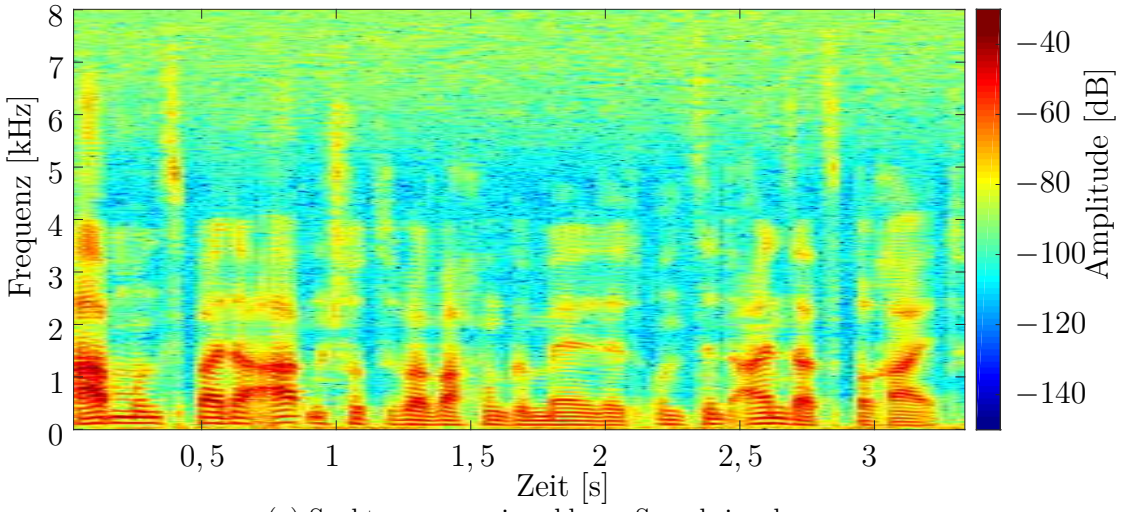
In diesem Kapitel wird eine abgeänderte Form des Exciters der Veröffentlichung von Herrn Bulling [Bul+16] genutzt, um die Sprachverständlichkeit bei der Kommunikationseinheit der Atemvollschutzmaske zu steigern. Der Exciter wird dabei beim Ohrhörer angewendet, so dass die taktischen Funksprüche und der Team-Funk verständlicher werden; dieses trägt zur Sicherheit im Einsatz bei. Die Sprachverständlichkeit in Verbindung mit der Atemvollschutzmaske ist so schlecht, da diese eine starke Dämpfungscharakteristik bis 500 Hz und ab 1400 Hz besitzt. Hinzu kommt die Nichtlinearität der Sprechmembran, welche bei sehr lauten Pegeln die Sprache verzerren kann. Ein Spektrogramm von Mikrofondaten mit und ohne Maske ist in der Abbildung 5.2 dargestellt, wobei es jeweils der

gleiche Sprecher mit dem gleichen gesprochen Satz war. Bei diesem Vergleich ist die vorher erwähnte Dämpfung zu sehen. Mit dem Exciter wird nun versucht, die Harmonischen über 1400 Hz wieder auszuprägen und damit die Sprachverständlichkeit zu steigern.

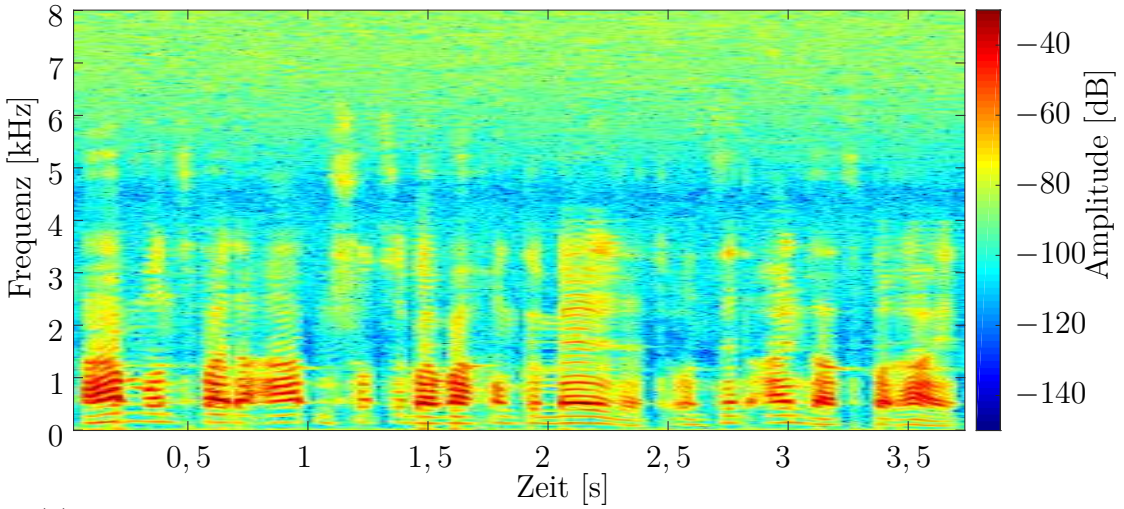
Der vereinfachte Signalflussgraph des Exciters ist in der Abbildung 5.3 zu sehen, wobei dieser ein Verzögerunselement z^{-d}, ein Verzerrungselement und ein Verstärkungselement g_h besitzt. In dem Verzerrerelement befinden sich ein Filter, nichtlineare Kennlinien und Gewichtungsfaktoren. Der Ausgang des Verzerrers wird mit g_h verstärkt auf das um die Laufzeit des Filters verzögerte Eingangssignal addiert. Somit ergibt sich das Ausgangssignal des Exciters zu

$$y_{ex}(n) = y_d(n) + g_h \cdot y_h(n). \tag{5.1}$$

In den folgenden zwei Unterkapiteln werden nun zwei Exciter vorgestellt, welche für ver-



(a) Spektrogramm eines klaren Sprachsignals



(b) Spektrogram eines mit der Kommunikatinseinheit aufgenommenen Sprachsignals

Abbildung 5.2: Vergleich von zwei Spektrogrammen; (a) klares Sprachsignal; (b) mit dem Mikrofon der Kommunikationseinheit aufgenommenes Sprachsignal.

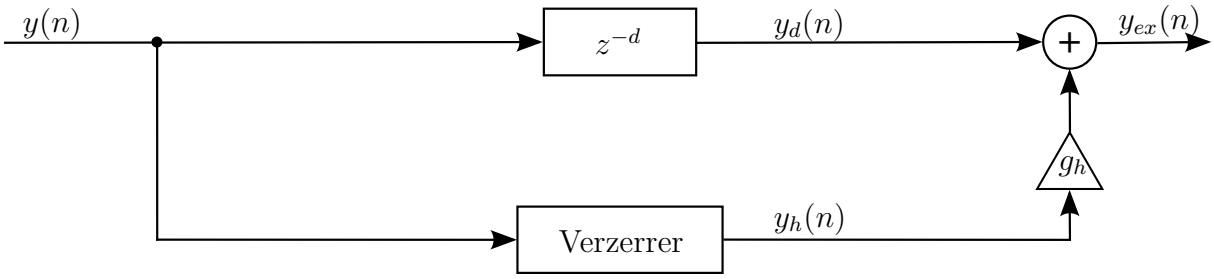


Abbildung 5.3: Übersicht des Exciter-Signalflussgraphen (verändert nach: [Bul+16])

schiedene Frequenzbereiche ausgelegt sind. Die Anwendung auf verschiedene Frequenzbereiche wird durch unterschiedliche nichtlineare Kennlinien erzielt.

5.1.1 Höhen-Exciter

Mit dem Höhen-Exciter sollen Harmonische im Bereich $f \geq 2\,\text{kHz}$ erzeugt werden, womit die starke Dämpfung der Atemvollschutzmaske (siehe Abb. 5.2 (b)) in den Harmonischen kompensiert werden soll. Somit soll die Sprachverständlichkeit in diesem Bereich damit gesteigert werden. Der Verzerrer der Abb. 5.3 wird hier genauer beschrieben, wobei dieser in einen Filter H_{hp}, zwei Pfade mit nichtlinearen Kennlinien und einen Gewichtungsfaktor aufgeteilt wird. Die Pfade mit den nichtlinearen Kennlinien sind K_u für die Erzeugung der ungeraden Harmonischen, K_g für die Erzeugung der geraden Harmonischen und ein Timbre-Faktor τ, welcher die geraden zu den ungeraden Harmonischen gewichtet. Der Faktor g_h gewichtet das verzerrte Signal und abschließend wird dies zu dem verzögertem Ausgangssignal addiert. Diese Komponenten sind zusammen in der Abbildung 5.4 abgebildet.

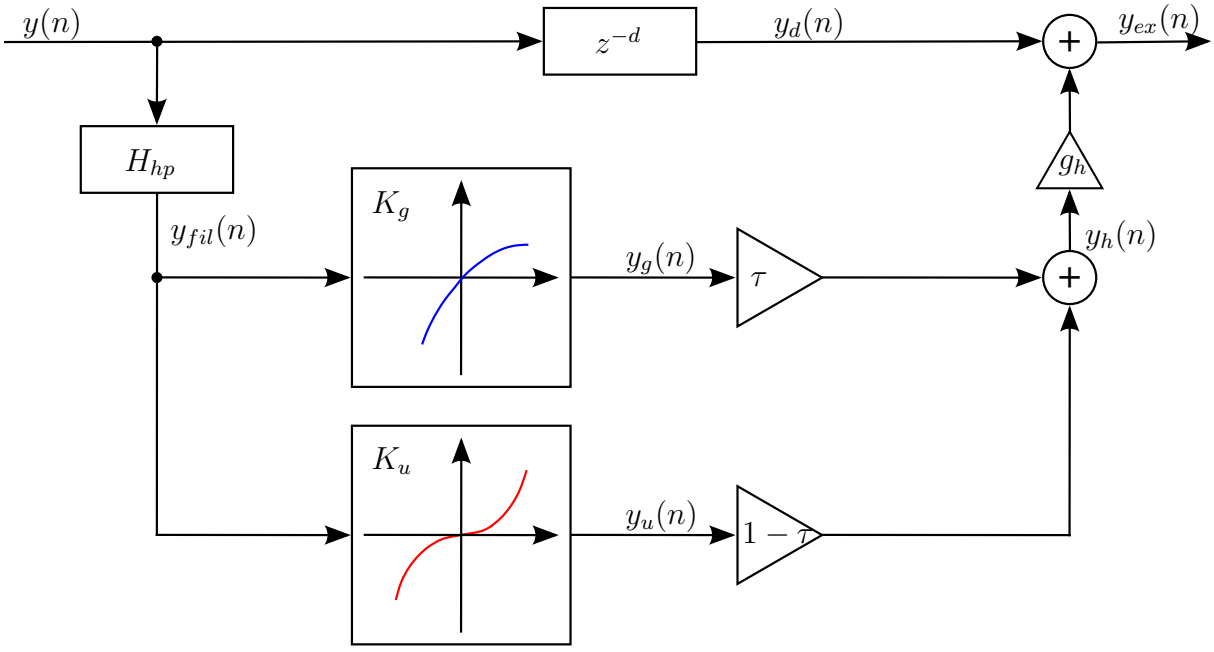


Abbildung 5.4: Vollständiger Exciter-Signalflussgraph (nach: [Bul+16])

Das Signal $y(n)$ ist der Ausgang der Mischung und Verstärkung. Das Signal wird daraufhin mit einem Hochpass mit einer 3-dB Grenzfrequenz von 1,6 kHz gefiltert. Diese soll tiefe Frequenzen mit hohen Pegeln dämpfen, so dass die Harmonischen hauptsächlich aus den höheren Frequenzen des Eingangs gebildet werden. Der Hochpass hat eine maximale Gruppenlaufzeit von 3 Abtastwerten, welches bei einer Abtastrate $f_s = 16\,\text{kHz}$ einer Zeit von 187,5 μs entspricht. Diese Laufzeit wird in dem Verzögerungpfad ausgeglichen, so dass $d = 3$ ist. Das hochpassgefilterte Signal $y_{fil}(n)$ wird nun durch die nichtlinearen Kennlinien K_g und K_u bearbeitet. Die Kennlinie K_g für die Erzeugung der geraden Harmonischen wird mit der Gleichung

$$K_g: \; y_g(n) = -0.5 \cdot y_{fil}(n)^2 + y_{fil}(n) \tag{5.2}$$

beschrieben und die Kennlinie K_u zur Erzeugung der ungeraden Harmonischen mit

$$K_u: \; y_u(n) = |y_{fil}(n)| \cdot y_{fil}(n)\ . \tag{5.3}$$

Diese nichtlinearen Kennlinien sind in der Abbildung 5.5 abgebildet, wobei die Kennlinie K aus der Addition der anderen beiden Kennlinien mit Berücksichtigung des Timbre-Faktors $\tau = 0,8$ resultiert.

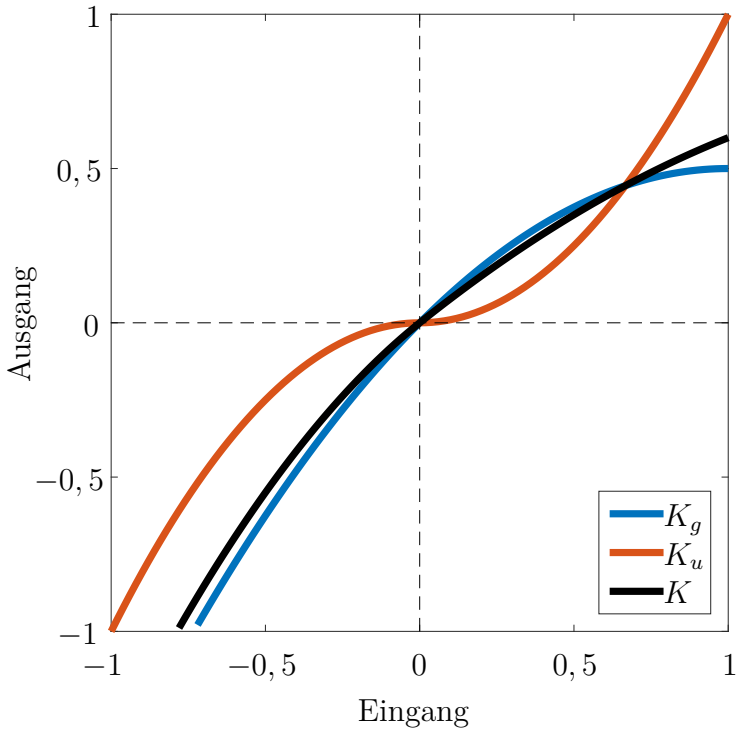


Abbildung 5.5: Kennlinie K als Addition der Kennlinie K_g für die Erzeugung der geraden Harmonischen und der Kennlinie K_u zur Erzeugung der ungeraden Harmonischen mit $\tau = 0,8$ für den Höhen-Exciter (nach: [Bul+16])

Der Timbre-Faktor ist zur Gewichtung der Klangfarbe, welche durch das Verhältnis der gerade zu den ungeraden Harmonischen erzielt wird. Der beste Wert des Timbre-Faktors ist mit 0,8 bewertet worden, welcher durch das subjektive Hörempfinden von 4

Probanden erzielt wurde. Die gewichteten geraden und ungeraden Harmonischen-Anteile werden daraufhin addiert und das Signal

$$y_h(n) = y_u(n) \cdot (1 - \tau) + y_g(n) \cdot \tau \tag{5.4}$$

ergibt sich. Dieses Signal wird daraufhin mit dem Faktor g_h multipliziert, wobei dieser Faktor maßgeblich zum Höreffekt des Exciters beiträgt. Dabei wurden die Werte $g_h = 8\,\text{dB}$, $g_h = 14\,\text{dB}$ und $g_h = 17,5\,\text{dB}$ genommen, welche in der subjektiven Betrachtung einen großen Unterschied aufweisen und dennoch einen guten Höreindruck zeigen. Mit diesen Verstärkungsfaktoren wurde ein Hörversuch durchgeführt, um bezogen auf des Zielsystem die optimale Parametrierung zu erkennen.

Subjektiver Hörtest

Beim Hörtest wurden zwei weibliche und zwei männliche Sprecher verwendet. Von diesen Sprechern wurde jeweils eine Version mit dem Exciter mit den genannten Verstärkungsfaktoren g_h erzeugt. Die Varianten und das Originalsignal der vier Sprecher wurden mit einem CMOS-ähnlichen Test verglichen [Grö+17]. Dabei wurden von jedem Sprecher alle Kombinationen der 4 Signale im Rahmen des Hörversuchs verglichen, also 6 Vergleiche pro Sprecher und insgesamt 24. Die Bewertung wurde gemäß der CMOS-Skala 4.6 durchgeführt. Der Hörtest wurde mit 10 Testpersonen durchgeführt, welche auf die Sprachverständlichkeit achten sollten und die Sprachqualität sollte nicht bewertet werden. Um die akustische Übertragung möglichst realistisch zu gestalten, wurde das Signal über den Ohrhörer einer Kommunikationseinheit wiedergegeben. Aus diesem Hörversuch resultiert das Ergebnis in Abb. 5.6 in Form eines Boxplots. Am besten wurde der Exci-

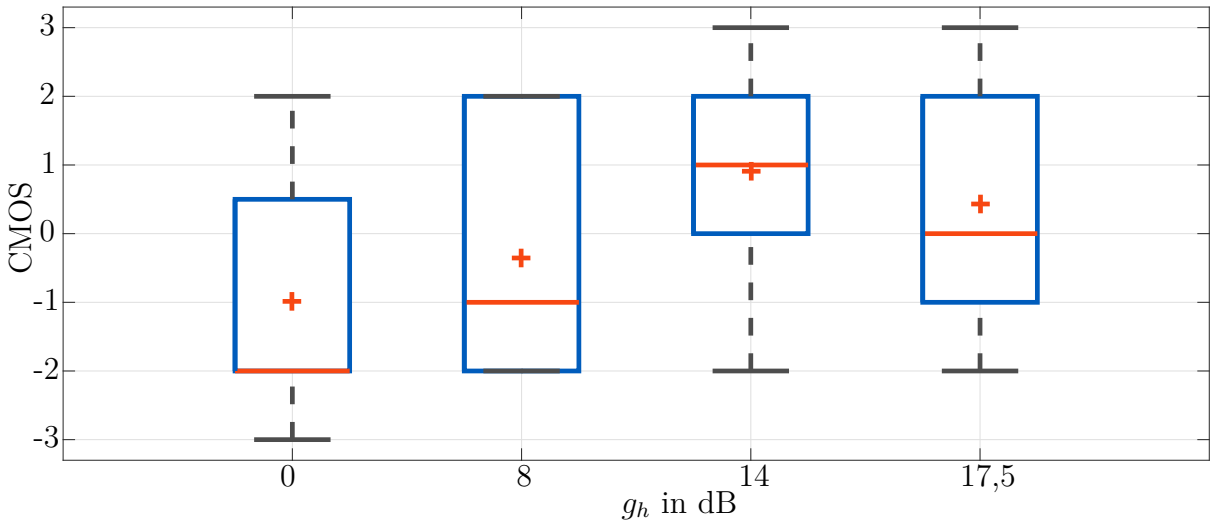


Abbildung 5.6: Darstellung der CMOS-Testergebnisse im Zusammenhang mit dem Höhen-Exciter auf der Bewertungsgrundlage von 10 Testpersonen bei der Bewertung von 4 Sprechern. Bewertet wurde primär die Verständlichkeit der Sprache und sekundär die Qualität der Sprache. Getestet wurden $g_h = 0\,\text{dB}$, $g_h = 8\,\text{dB}$, $g_h = 14\,\text{dB}$ und $g_h = 17,5\,\text{dB}$

ter mit dem Verstärkungsfaktor von $g_h = 14\,\text{dB}$ bewertet, gefolgt vom Verstärkungsfaktor

$g_h = 17,5\,\text{dB}$. Daraufhin folgt der Verstärkungsfaktor von $g_h = 8\,\text{dB}$, der Verstärkungsfaktor von $g_h = 0\,\text{dB}$ wurde am schlechtesten bewertet. Die Varianz bezogen auf die Quartile (beinhaltet 50 %) ist bei der Bewertung des Exciters mit dem Verstärkungsfaktor von $g_h = 14\,\text{dB}$ am geringsten und der Mittelwert und der Median sind am höchsten. Alle diese Kriterien zeigen, dass der Exciter mit dem Verstärkungsfaktor $g_h = 14\,\text{dB}$ im Vergleich die beste Bewertung erhielt. Somit kann mit diesem Exciter die Sprachverständlichkeit gesteigert werden. Dieser Exciter ist in der Abbildung 5.7 als Spektrogramm dargestellt, wobei im oberen Plot das originale Mikrofonsignal und im unteren Plot das Mikrofonsignal bearbeitet mit dem Höhen-Exciter mit den Parametern $\tau = 0,8$ und $g_h = 14\,\text{dB}$ dargestellt ist. Hier ist zu erkennen, dass die Harmonischen deutlich stärker ausgeprägt sind, welches zur Steigerung der Sprachverständlichkeit führt. Zusätzlich ist zu erkennen, dass der Rauschpegel ab 5 kHz deutlich verstärkt wurde. Damit dieses nicht störend ist, wird dieses mittels einer Geräuschunterdrückung 4.4 gedämpft. Der Höhen-Exciter wird in der Kommunikationseinheit aufgrund der Verbesserung der Sprachverständlichkeit verwendet.

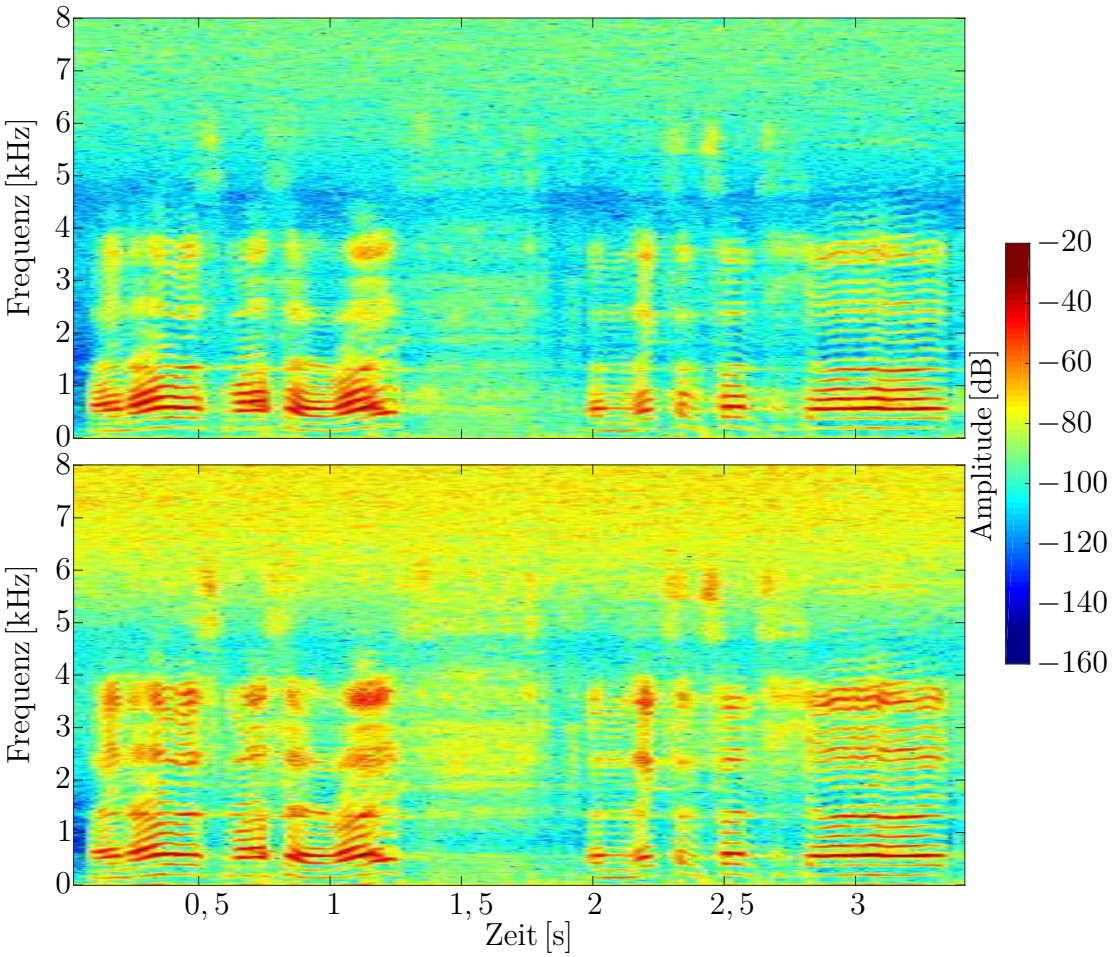


Abbildung 5.7: Signal nach Bearbeitung mit dem Höhen-Exciter mit $\tau = 0,8$ und $g_h = 14\,\text{dB}$; Eingangssignal y (oberer Plot); Signal nach Bearbeitung mit dem Höhen-Exciter y_{ex} (unterer Plot).

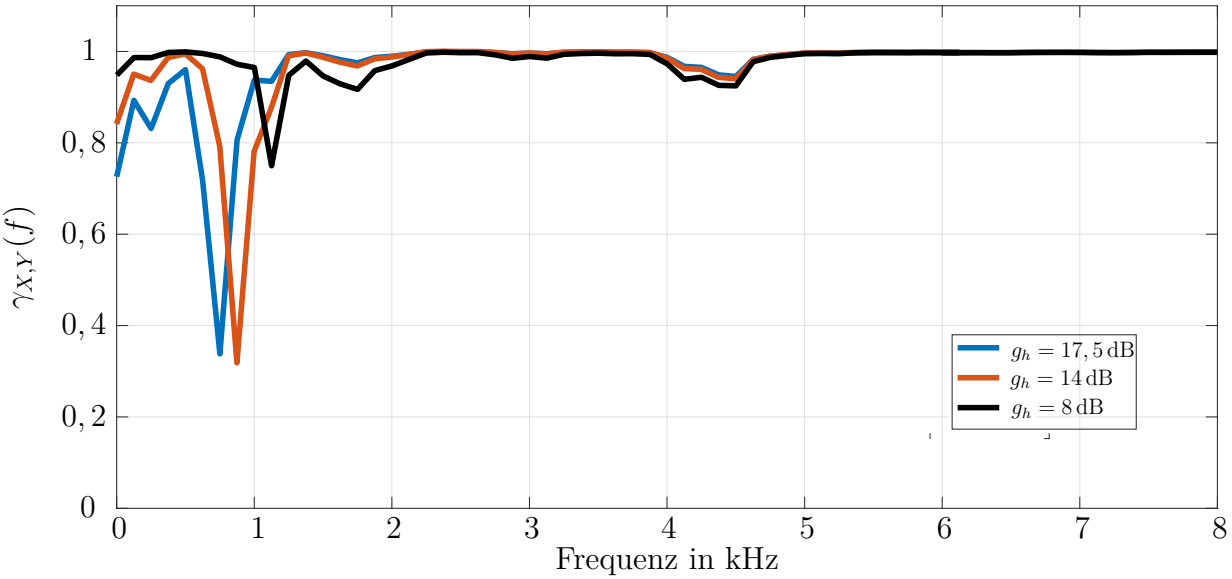


Abbildung 5.8: Kohärenzanalyse bei einem Höhen-Exciter mit $g_h = 8\,\mathrm{dB}$, $g_h = 14\,\mathrm{dB}$, $g_h = 17,5\,\mathrm{dB}$

Kohärenzanalyse des Höhen-Exciters

Um den Einfluss des Exciters auf den Rückkopplungskompensator zu untersuchen, wurde ermittelt, wie stark das Signal durch den Höhen-Exciter mit den Parametern $\tau = 0,8$ und $g_h = 8\,\mathrm{dB}$, $g_h = 14\,\mathrm{dB}$ und $g_h = 17,5\,\mathrm{dB}$ dekorreliert wurde. Je stärker das Signal dekorreliert ist, je besser können auftretende Rückkopplungen vom Rückkopplungskompensator entfernt werden. Dafür wurde analog zu der Analyse der Dekorrelation des Frequenzversatzes aus Kapitel 4.3.4 eine Kohärenzanalyse durchgeführt. Diese ist in der Abbildung 5.8 abgebildet. Dabei ist zu sehen, dass eine Dekorrelation hauptsächlich im Bereich um 1 kHz auftritt; je größer der Verstärkungsfaktor g_h gewählt wurde, desto größer ist die Dekorrelation in diesem Bereich. In den anderen Frequenzbereichen ist keine signifikante Dekorrelation zu beobachten.

5.1.2 Mitten-Exciter

Bei dem Mitten-Exciter sollen Harmonische zwischen 1,4 kHz und 2 kHz erzeugt werden, so dass die Dämpfung der Maske über 1,4 kHz ausgeglichen wird. Der Signalfluss des Mitten-Exciters ist ähnlich wie der Höhen-Exciter aufgebaut, wobei sich der Verzerrungspfad unterscheidet. In Abbildung 5.9 ist der gesamte Signalflussgraph dargestellt. Bei diesem

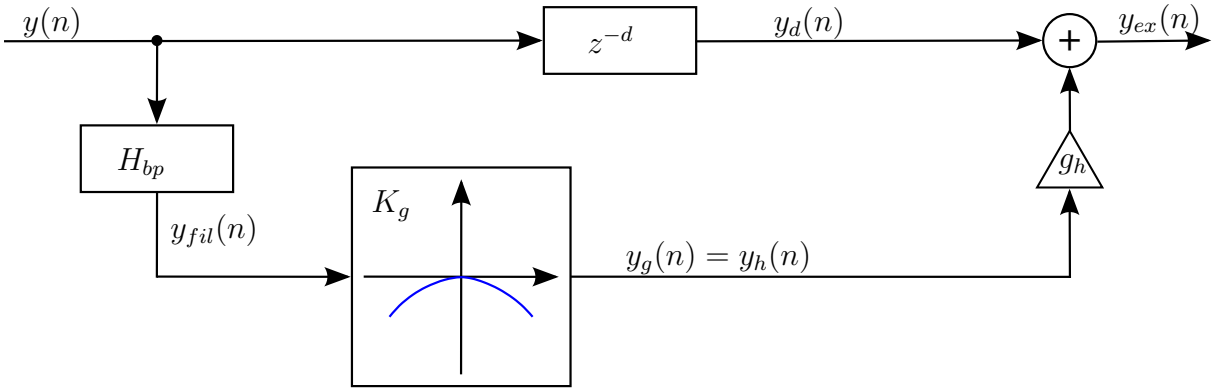


Abbildung 5.9: Signalflussgraph des Mitten-Exciters

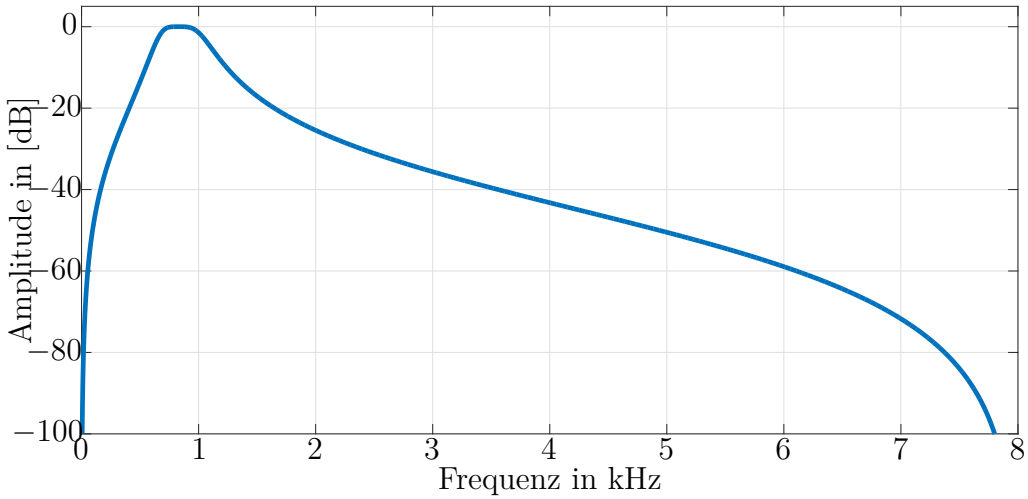


Abbildung 5.10: Frequenzverlauf des IIR-Butterworth-Bandpasses $H_{ex,mid}$

Verzerrungspfad wird das Eingangssignal durch einen IIR-Bandpassfilter H_{bp} 4. Ordnung mit den 3-dB Grenzfrequenzen von 650 Hz und 1050 Hz gefiltert und ist in Abbildung 5.10 abgebildet. Dieses Filter hat eine Gruppenlaufzeit von 1,6 ms, welches bei der Abtastrate von $f_s = 16$ kHz ca. $d = 25$ Abtastwerten entspricht, womit das Signal $y(n)$ verzögert wird. Daraufhin wird die nichtlineare Kennlinie K_g auf das gefilterte Signal $y_{fil}(n)$ angewendet, welche in Abbildung 5.11 abgebildet ist und durch

$$K_g: \; y_g(n) = -0.5 \cdot y_{fil}^2(n) \tag{5.5}$$

berechnet wird. Das Ausgangssignal des Exciters wird durch

$$y_{ex}(n) = y_d(n) + g_h \cdot y_h(n) \tag{5.6}$$

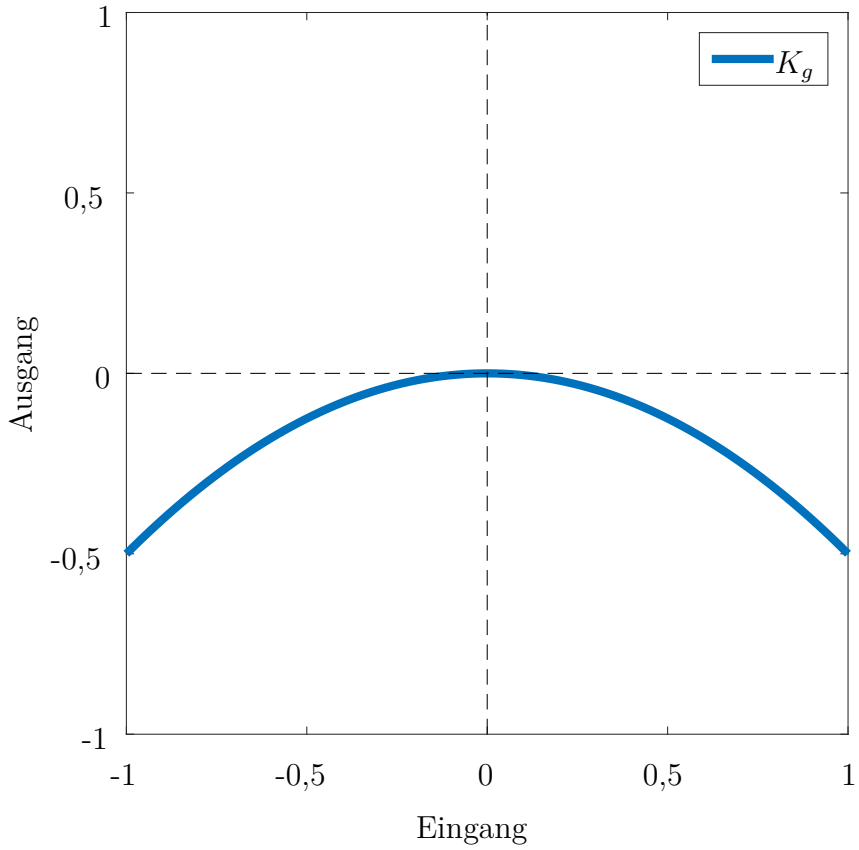


Abbildung 5.11: Kennlinie K_g für die Erzeugung der geraden Harmonischen für den Mitten-Exciter

erzeugt, wobei g_h der Verstärkungsfaktor des verzerrten Signals ist. Für die Wahl des Verstärkungsfaktors g_h wurde genau wie bei dem Höhen-Exciter ein Hörversuch mit den Verstärkungsfaktoren $g_h = 8\,\text{dB}$, $g_h = 14\,\text{dB}$ und $g_h = 17,5\,\text{dB}$ zur Optimierung der Parametrierung vorgenommen.

Subjektiver Hörtest

Der Hörtest wird analog zu dem Hörtest des Höhen-Exciters aus Kapitel 5.1.1 durchgeführt. Es wurden wieder bei jedem der 4 Sprecher alle Kombination der Verstärkungsfaktoren $g_h = 0\,\text{dB}$, $g_h = 8\,\text{dB}$, $g_h = 14\,\text{dB}$ und $g_h = 17,5\,\text{dB}$ verglichen und gemäß der C-MOS-Skala bewertet [Grö+17]. Der Hörtest wurde mit 10 Testpersonen durchgeführt, welche auf die Sprachverständlichkeit achten sollten, die Sprachqualität sollte nicht bewertet werden; die Signale wurden auf einem Ohrhörer einer Kommunikationseinheit abgespielt. Das Ergebnis dieses Hörtests ist in der Abbildung 5.12 in Form eines Boxplots angegeben. Bei diesem Testergebnis ist zu sehen, dass bei zunehmendem Verstärkungsfaktor die Bewertung schlechter wird, wobei sich aber die Verstärkungsfaktoren $g_h = 0\,\text{dB}$, $g_h = 8\,\text{dB}$ und $g_h = 14\,\text{dB}$ in der Bewertung nicht signifikant unterscheiden und das Signal mit dem Verstärkungsfaktor $g_h = 17,5\,\text{dB}$ signifikant schlechter ist. Somit ist mit dem Mitten-Exciter bezogen auf die Sprachverständlichkeit keine Verbesserung erzielbar und je höher der Verstärkungsfaktor gewählt wird, desto schlechter wird die Sprachver-

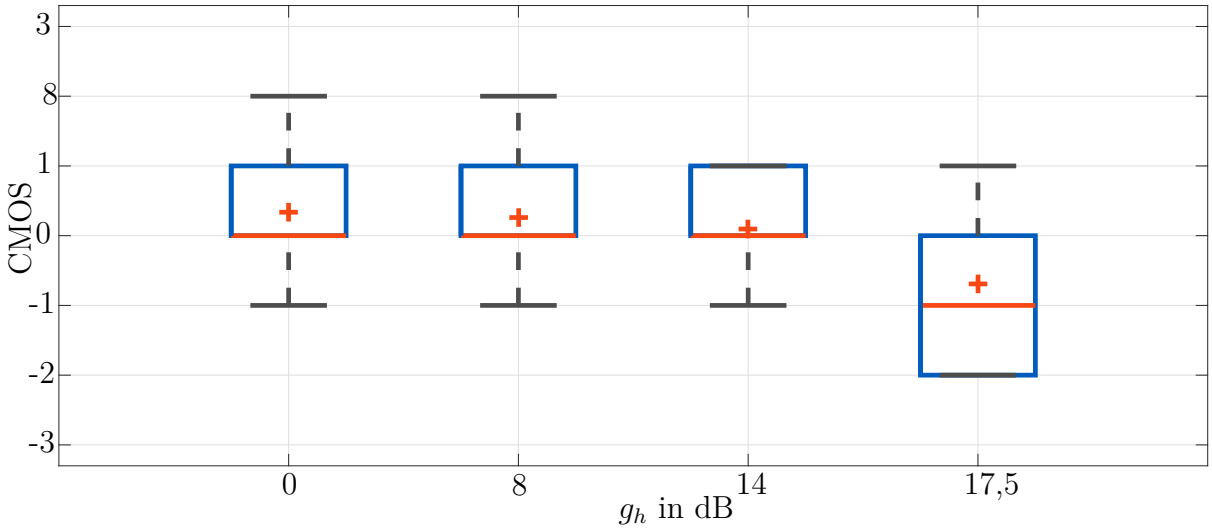


Abbildung 5.12: Darstellung der CMOS-Testergebnisse im Zusammenhang mit dem Mitten-Exciter auf der Bewertungsgrundlage von 10 Testpersonen bei der Bewertung von 4 Sprechern; bewertet wurde primär die Verständlichkeit der Sprache und sekundär die Qualität der Sprache; getestet wurden $g_h = 0\,\text{dB}$, $g_h = 8\,\text{dB}$, $g_h = 14\,\text{dB}$ und $g_h = 17,5\,\text{dB}$

ständlichkeit. In der Abbildung 5.13 ist ein Spektrogramm mit dem Originalsignal im oberen Plot und im unteren Plot das verzerrte Signal y_{ex} mit dem Verstärkungsfaktor von $g_h = 17,5\,\text{dB}$ abgebildet. Hier ist zu sehen, dass sich die Harmonischen im Bereich zwischen 1,4 kHz und 2 kHz ausprägen. Der Mitten-Exciter wird in der Kommunikationseinheit nicht verwendet, da dieser die Sprachverständlichkeit nicht signifikant steigern kann.

Kohärenzanalyse des Mitten-Exciters

Um den Einfluss des Exciters auf den Rückkopplungskompensator zu untersuchen, wurde ermittelt, wie stark das Signal durch den Mitten-Exciter dekorreliert wird. Dafür wurde von den verzerrten Signalen mit den Verstärkungsfaktoren $g_h = 8\,\text{dB}$, $g_h = 14\,\text{dB}$ und $g_h = 17,5\,\text{dB}$ eine Kohärenzanalyse durchgeführt und in Abbildung 5.14 dargestellt. Bei allen Verstärkungsfaktoren wird das Signal im Bereich von ca. 1,4 kHz bis 2 kHz dekorreliert, so dass die Kohärenz in diesen Bereichen einen Wert von ca. 0,2 annimmt und sich die verschiedenen Verstärkungsfaktoren nicht signifikant unterscheiden. In den restlichen Frequenzbereichen ist keine Dekorrelation zu erkennen, welches widerspiegelt, dass der Mitten-Exciter nur Harmonische in diesem Bereich ausprägt.

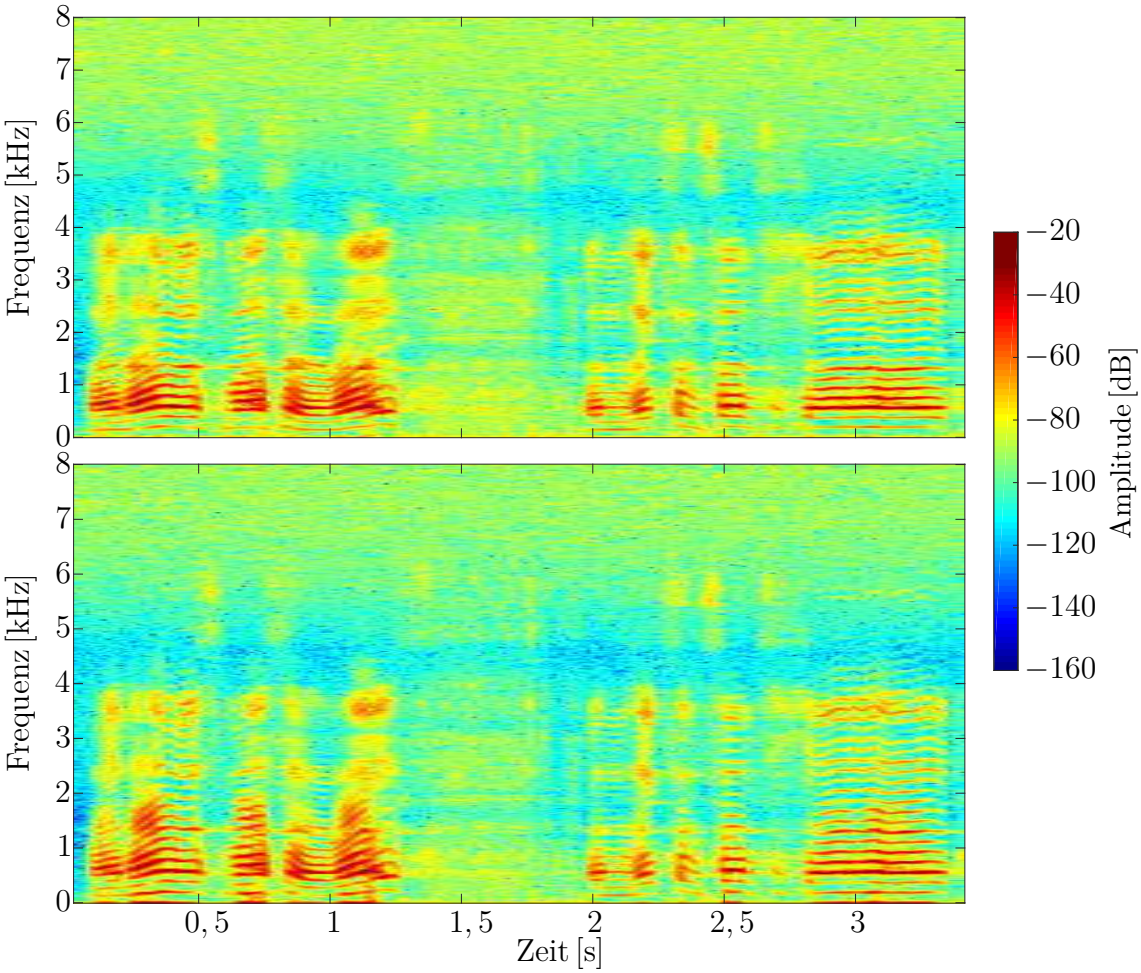


Abbildung 5.13: Originalsignal y (oberer Plot); Signal y_{ex} nach Bearbeitung mit dem Mitten-Exciter ($g_h = 17,5\,\mathrm{dB}$) (unterer Plot).

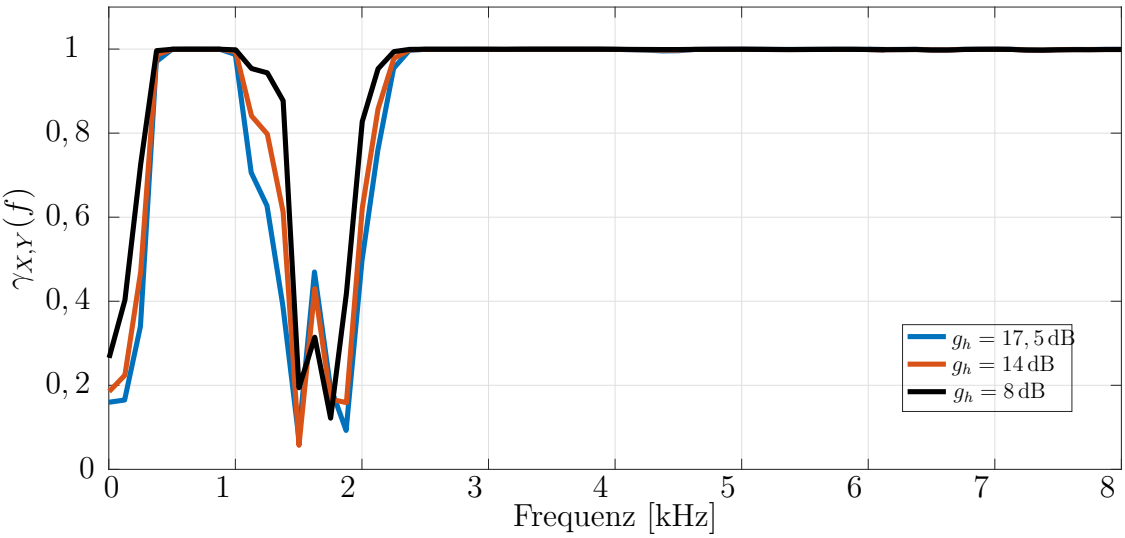


Abbildung 5.14: Kohärenzanalyse bei einem Mitten-Exciter mit $g_h = 8\,\mathrm{dB}$, $g_h = 14\,\mathrm{dB}$, $g_h = 17,5\,\mathrm{dB}$

5.2 Equalizer-Filter

Die VA-Lautsprecher der Kommunikationseinheit der Atemvollschutzmaske müssen aufgrund des Formfaktors in sehr kleine Gehäuse passen und durch das Einsatzszenario Wärme, Kälte und Feuchtigkeit aushalten. Die verwendeten Lautsprecher in der Kommunikationseinheit können alle Kriterien erfüllen, allerdings haben diese Lautsprecher unterhalb von 800 Hz eine sehr starke Dämpfung. In Verbindung mit der Dämpfung der Atemvollschutzmaske sind zusätzlich im Anregungssignal alle Frequenzen über 1,4 kHz stark gedämpft und somit können mit dem VA-Lautsprecher die gedämpften Frequenzen wieder verstärkt werden. Zum Abstimmen des Frequenzgangs dient ein Equalizer-Filter [Grü08, S. 74 ff], wobei diese Abstimmung beispielsweise durch Shelving-, Peak-, Notch-, Bandpass-, Tiefpass- und/oder Hochpass-Filter geschieht. Bei den einzelnen Filtern sind verschiedene individuelle Parameter notwendig zur Bestimmung des Frequenzgangs, welche beispielsweise die 3dB-Grenzfrequenz, die Verstärkung, die Mittenfrequenz und die Bandbreite sein können, wobei bei jedem Filter-Typ noch dessen Filterordnung wichtig ist, da damit beispielsweise die Flankensteilheit verändert werden kann. Diese Equalizer-Filter können mit einem FIR-Filter oder einem IIR-Filter umgesetzt werden. Das IIR-Filter benötigt für die gleiche Filterung eine geringere Filterordnung als das FIR-Filter, wodurch die Gruppenlaufzeit bei dem IIR-Filter geringer ist. Die Gruppenlaufzeit kann beim IIR-Filter nicht konstant sein, welches bei der Wahl betrachtet werden muss. Die Vorteile des FIR-Filters liegen in der garantierten Stabilität [Grü08, S. 12], welche bei dem IIR-Filter durch das Design und die Implementierung sichergestellt werden muss. In Bezug auf die Realisierung des Equalizer-Filters wird ein IIR-Filter 2. Ordnung in der Direkt-Form 2 verwendet [Grü08; Lai03; Kuo+01], welches auch Biquad-IIR-Filter genannt wird, da sowohl die Rechenleistung gespart werden muss als auch ebenfalls die Gruppenlaufzeit wichtig ist. Das verwendete Biquad-IIR-Filter in der Direkt-Form 2 ist in der Abbildung 8.2 dargestellt, wobei diese Abbildung durch die Differenzengleichung

$$y_{\mathrm{eq}}(n) = \sum_{i=1}^{2} a_i \cdot y_{\mathrm{ex}}(n-i) + \sum_{i=0}^{2} b_i \cdot x(n-i) \tag{5.7}$$

beschrieben wird. Die zugehörige Übertragungsfunktion lautet

$$H(z) = \frac{b_0 + b_1 z^{-1} + b_2 z^{-2}}{1 - a_1 z^{-1} - a_2 z^{-2}}, \tag{5.8}$$

womit die Pol- und Nullstellen bestimmt werden können [Zöl+02]. Die Direkt-Form 2 kann bei der Verwendung mehrerer Filter kaskadiert werden, welche *Second Order Structure*-IIR-Filter (SOS-IIR-Filter) bezeichnet werden [Hoc07, S. 11][Wel+16, S. 215 ff]. Bei der Kaskadierung wirkt sich der Quantisierungsfehler geringer als bei Verwendung vieler einzelner Filter aus, welches für die Festkomma-Implementierung sehr wichtig ist. Die Übertragungsfunktion ergibt sich in der kaskadierten Form zu

$$H(z) = g \cdot \frac{b_0 + b_1 z^{-1} + b_2 z^{-2}}{1 - a_1 z^{-1} - a_2 z^{-2}}, \tag{5.9}$$

wobei g der Verstärkungsfaktor ist. Bei der Festkomma-Implementierung müssen die Koeffizienten so skaliert werden, dass bei der Filterung keine Überläufe geschehen und trotzdem die Genauigkeit sehr hoch ist. Die a-Koeffizienten können nicht skaliert werden, da a_0

immer 1 ist und bei der Skalierung von a_1 und a_2 würden sich die Pole relativ zu einander verschieben [Hoc07]. Daher wird das Eingangssignal so skaliert, dass bei der Berechnung mit den a-Koeffizienten kein Überlauf geschehen kann und diese Skalierung wird nach der letzten Kaskade wieder zurück skaliert. Die *b*-Koeffizienten können ohne Einschränkung in das gewünschte Zahlenformat skaliert werden, wodurch das Format des Zwischenresultats und des Ausgangssignals nicht überläuft.

5.2.1 Verwendete Equalizer

Bei der Übertragung des Mikrofonsignals zum Funkgerät wird der Frequenzgang mit dem Equalizer aus der Abbildung 5.15 bearbeitet. Dieses ist ein *Shelving*-Filter, welcher bis 2,5 kHz maximal 7 dB dämpft und ab 2,5 kHz maximal 2 dB verstärkt.

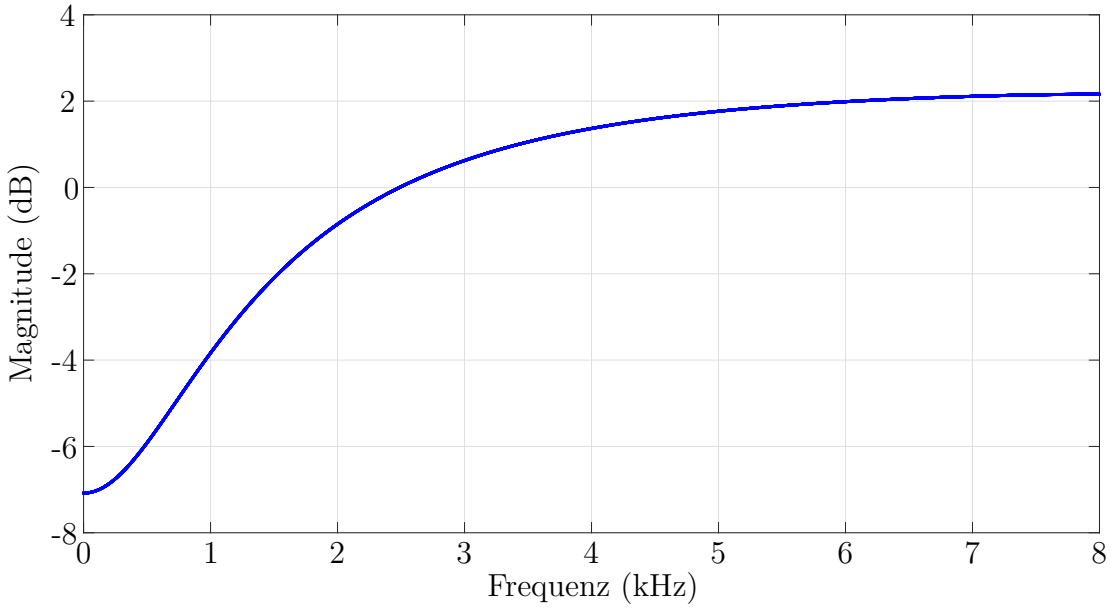


Abbildung 5.15: Frequenzgang des Equalizer vom taktischen Funkausgangssignal

Für die VA-Lautsprecher sind die Equalizer sehr wichtig, da somit der Lautsprecher sehr gut abgestimmt werden kann. Das hier verwendete Filter hat zwei Hochpässe mit einer 3 dB-Grenzfrequenz von 1,2 kHz, so dass der Lautsprecher keine tieffrequenten Anteile ausgibt und nur die Dämpfung der Maske in höheren Frequenzen ausgleicht. Zusätzlich werden ein *Shelving*-Filter und zwei *Notch*-Filter genutzt. Das *Shelving*-Filter verstärkt das Signal ab 5 kHz um maximal 8 dB. Die *Notch*-Filter haben eine zentrale Frequenz bei 5 kHz und der zweite bei 6,2 kHz und diese werden genutzt, um die sehr kritischen Rückkopplungsfrequenzen zu dämpfen. Die Kombination dieser Filter ist in der Abbildung 5.16 dargestellt.

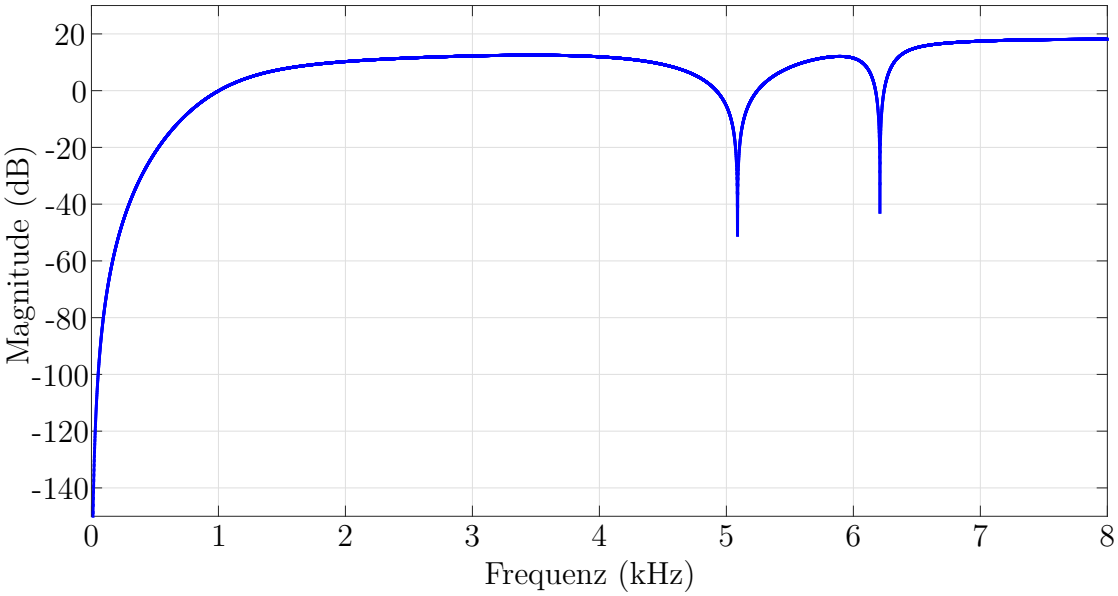


Abbildung 5.16: Frequenzgang des Equalizer vom VA-Signal

5.3 Regelverstärker

Der verwendete Regelverstärker wurde 2019 im *EURASIP Journal on Audio, Speech, and Music Processing* veröffentlicht [Bro+19]. Die Dynamik von Audiosignalen ist in vielen Systemen und Anwendungssituationen sehr unterschiedlich, daher ist es in einigen Systemen erforderlich die Dynamik zu beeinflussen und anzupassen [DG+12; Zöl+02; Zöl13; Dic+08]. Diese Dynamikanpassung kann aus verschiedenen Gründen erforderlich sein, dies kann beispielsweise im Auto die Anpassung der Dynamik an die Innengeräusche oder bei der Kommunikationseinheit der Atemvollschutzmaske die Dynamikanpassung der Ohrhörer an das eingehende Signal sein. Bei der Dynamikanpassung wird in fünf verschiedene Wirkungsformen unterschieden:

- der *Limiter* begrenzt das Signal, so dass beispielsweise vor Übersteuerung oder Gehörschäden geschützt wird;
- der Kompressor komprimiert das Signal, so dass große Eingangspegeländerungen kleinere Ausgangspegeländerungen hervorruft;
- der Durchleiter gibt das Eingangssignal in Bezug auf die Dynamik unverändert wieder, so dass der Verstärkungsfaktor in dem Bereich des Durchleiters konstant ist;
- der Expander vergrößert den Ausgangspegelbereich in Bezug auf den Eingangspegelbereich, so dass kleine Änderungen des Eingangspegels größere des Ausgangspegels hervorrufen und
- das *Noisegate* bewirkt eine Rauschunterdrückung.

Durch den Pegel des Eingangssignals wird der Dynamikbereich und durch das Verhältnis des Eingangspegels zu dem gewünschtem Ausgangspegel der Verstärkungsfaktor bestimmt. Der gewünschte Ausgangspegel wird typischerweise durch eine Dynamik-Kennlinie

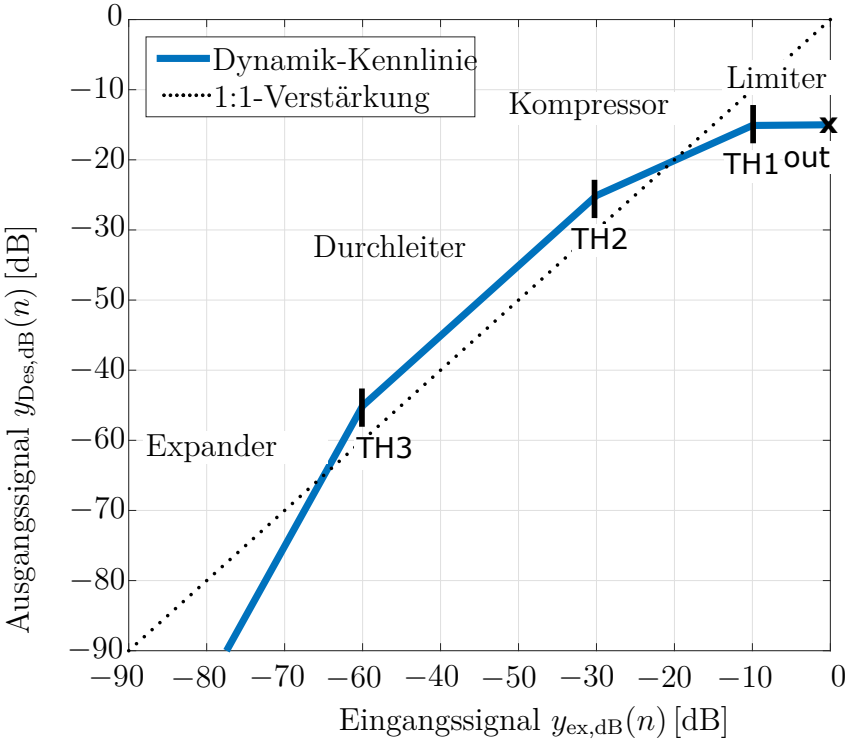


Abbildung 5.17: Beispiel einer Dynamik-Kennlinie

abgebildet, welche beispielhaft in Abbildung 5.17 dargestellt ist. Hier sind die verschiedenen Dynamik-Bereiche jeweils gekennzeichnet und auf der X-Achse das Eingangssignal in dB und auf Y-Achse das Ausgangssignal in dB dargestellt. Die gestrichelte Linie stellt eine 1:1-Verstärkung dar, bei dessen Anwendung keine Dynamikanpassung geschieht. Ein Problem bei den Übergängen zwischen den einzelnen Dynamikbereichen sind die harten Übergänge zwischen den Geraden. Diese werden auch als *hard knee* bezeichnet. Dieses *hard knee* führt zu einem unnatürlichem Klang bei einem Übergang zwischen den Geraden, weshalb für einen natürlichem Klang ein weicher Übergang zwischen den Geraden, ein sogenanntes *soft knee*, verwendet wird. Ein beispielhaftes *soft knee* ist in der Abbildung 5.18 dargestellt. Dieses *soft knee* ist in der Umsetzung durch eine lineare Approximation mit Hilfe mehrerer Zwischengeraden realisiert, wobei der Unterschied zwischen einer guten linearen Approximation und einem wirklichen *soft knee* nicht zu hören ist.

Die Berechnung des Verstärkungsfaktors $g_{\mathrm{dyn}}(n)$ wird mit Hilfe der Dynamik-Kennlinie bestimmt, wobei allerdings noch mehr Verarbeitungsschritte notwendig sind. Hierzu ist der Signalflussgraph der Dynamikanpassung in der Abbildung 5.19 zu sehen. Von dem Eingangssignal y_{eq} wird eine Pegelschätzung vollzogen, wobei bei der Pegelschätzung als erstes der Betrag des Eingangssignal $|y_{\mathrm{eq}}(n)|$ gebildet und daraufhin durch eine Glättung mit einem IIR-Filter erster Ordnung eine Pegelglättung vorgenommen wird:

$$\left|\overline{y_{\mathrm{eq}}}(n-1)\right| = \alpha_{\mathrm{sm}}(n) \cdot \left|y_{\mathrm{eq}}(n)\right| + (1-\alpha_{\mathrm{sm}}(n)) \cdot \left|\overline{y_{\mathrm{eq}}}(n-1)\right|. \tag{5.10}$$

Die Glättungskonstante $\alpha_{\mathrm{sm}}(n)$ wird durch die Ansprechzeit α_{attack} bestimmt, wenn der Betrag $\left|y_{\mathrm{eq}}(n)\right|$ größer als der vorher geschätzte Pegel $\left|\overline{x}(n-1)\right|$ ist. Ansonsten wird die

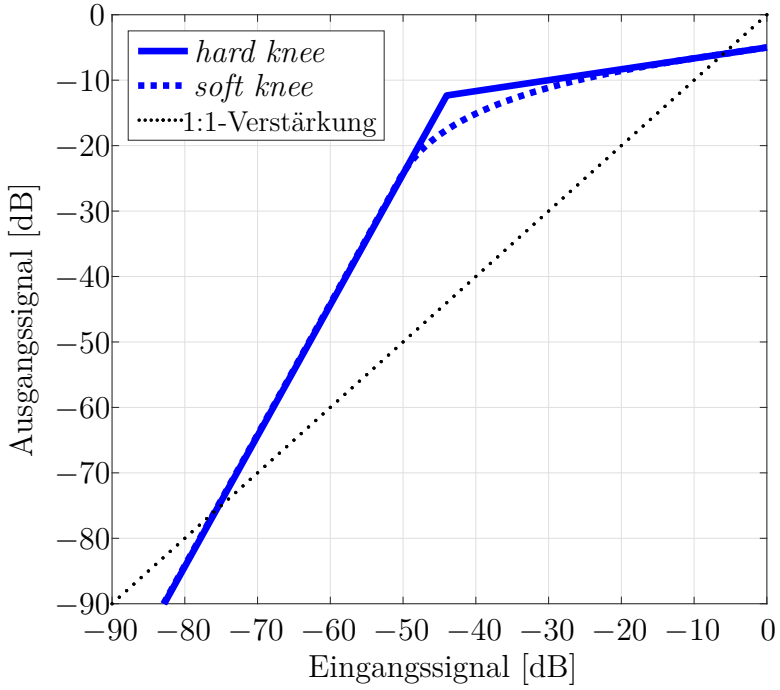


Abbildung 5.18: Beispiel einer Dynamik-Kennlinie mit *hard knee*- und *soft knee*-Übergang

Glättungskonstante durch die Rücklaufzeit α_{release} bestimmt, wodurch sich

$$\alpha_{\text{sm}}(n) = \begin{cases} \alpha_{\text{attack}} & \text{, wenn } \left|y_{\text{eq}}(n)\right| > \left|\overline{y_{\text{eq}}}(n-1)\right|, \\ \alpha_{\text{release}} & \text{, sonst} \end{cases} \tag{5.11}$$

ergibt. Die Ansprechzeit ist hier schneller als die Rücklaufzeit, so dass die Verstärkung $g_{\text{dyn}}(n)$ nicht zu stark schwankt. Die Ansprechzeit ist in der Implementierung mit $\alpha_{\text{attack}} \hat{=} 10\,\text{ms}$ und die Rücklaufzeit mit $\alpha_{\text{release}} \hat{=} 200\,\text{ms}$ gewählt worden. Nach der Pegelschätzung wird

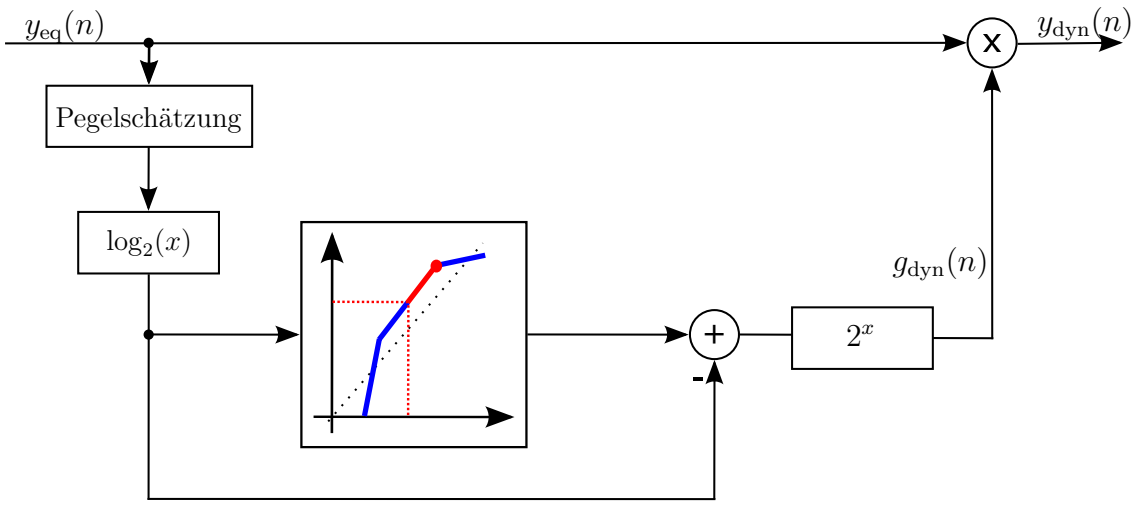


Abbildung 5.19: Signalflussgraph des Regelverstärkers

der Logarithmus des Signals gebildet, wobei wiederum der Logarithmus Dualis wie im Kapitel 4.2.2 angewandt wird, so dass sich der Aufwand bei der Berechnung reduziert. Es resultiert das Signal $y_{\mathrm{eq,dB}}(n)$ in dB, welches das Eingangssignal der Dynamik-Kennlinie ist. Mit $y_{\mathrm{eq,dB}}(n)$ und der Dynamik-Kennlinie wird der gewünschte Ausgangspegel $y_{\mathrm{Des,dB}}(n)$ ermittelt und damit der logarithmische Verstärkungsfaktor

$$g_{\mathrm{dyn,dB}}(n) = y_{\mathrm{Des,dB}}(n) - y_{\mathrm{eq,dB}}(n) \tag{5.12}$$

berechnet. Der lineare Verstärkungsfaktor

$$g_{\mathrm{dyn}}(n) = 2^{g_{\mathrm{dyn,dB}}(n)} \tag{5.13}$$

wird durch die Transformation in den linearen Bereich erhalten. Dieser Verstärkungsfaktor wird mit dem Eingangssignal multipliziert, so dass sich das Ausgangssignal

$$y_{\mathrm{dyn}}(n) = y_{\mathrm{eq}}(n) \cdot g_{\mathrm{dyn}}(n) \tag{5.14}$$

ergibt.

5.3.1 Verwendete Dynamik-Kennlinien

Die Dynamikanpassung wird für die Ohrlautsprecher und die VA-Lautsprecher genutzt. Bei den Ohrlautsprechern wird die Dynamik-Kennlinie aus der Abbildung 5.20 verwendet, wobei die maximale Aussteuerung des Ausgangs bei -5 dB liegt, wenn das Eingangssignal einen Pegel von 0 dB aufweist. Es werden zwei verschiedene Kompressor-Bereiche genutzt, wobei der erste vom Eingangspegel aus gesehen bei 0 dB startet und bei -20 dB endet. Dieser Bereich hat eine Steigung von 1/7, so dass bei dem Endpunkt beim Anregungspegel von -20 dB ein Ausgangspegel von -7.857 dB erzeugt wird. Ab dem Endpunkt dieser Kompressor-Geraden wird das *soft knee* als Übergang genutzt, so dass der zweite Kompressor-Bereich ab -25 dB anfängt. Diese Kompressor-Gerade hat eine Steigung 1/1,3, so dass alle Eingangssignale verstärkt werden. Dies ist notwendig um die geforderte Lautstärke in Verbindung mit der Hardware zu erreichen, so dass die Funksprüche immer deutlich verstanden werden können.
Bei den VA-Lautsprechern wird die Dynamik-Kennlinie aus der Abbildung 5.21 verwendet, wobei die maximale Aussteuerung des Ausgang bei -10 dB liegt, wenn das Eingangssignal einen Pegel von 0 dB aufweist. Es werden ein Kompressor-Bereiche und ein Durchleiter-Bereich genutzt, wobei der Kompressor-Bereich gesehen vom Eingangspegel bei 0 dB startet und bei -17 dB endet. Dieser Bereich hat eine Steigung von 3/8, so dass bei dem Endpunkt beim Anregungspegel von -17 dB ein Ausgangspegel von -16.38 dB erzeugt wird. Ab dem Endpunkt dieser Kompressor-Geraden wird das *soft knee* als Übergang genutzt, so dass der Durchleiter-Bereich ab -21 dB anfängt. Somit werden im Bereich der Durchleiter-Geraden alle Eingangssignale um die Parallelverschiebung der Winkelhalbierenden verstärkt, welches in diesem Fall einem Verstärkungsfaktor von 1,87 dB entspricht. Mit dieser gesamten Dynamik-Kennlinie werden die VA-Lautsprecher im Kompressor-Bereich dynamisch begrenzt und die restlichen Signale werden leicht verstärkt.

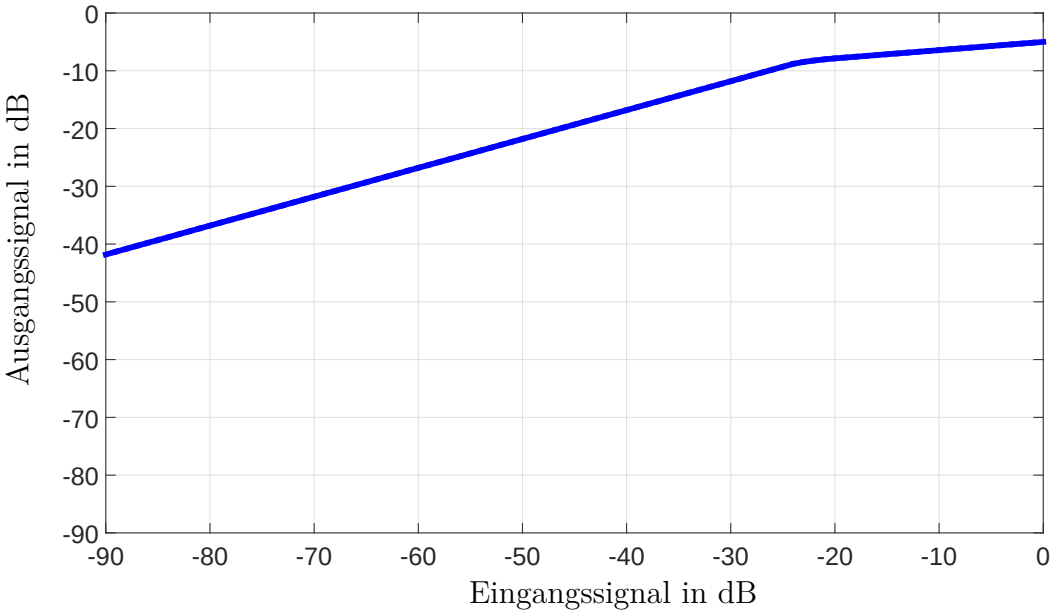


Abbildung 5.20: Dynamik-Kennline der Ohrlautsprecher

5.3.2 Anwendung der Dynamikanpassung am Beispiel der VA-Kennlinie

Für die VA-Lautsprecher wird die Dynamikkennlinie aus der Abbildung 5.21 verwendet, welche einen Kompressor-Bereich und einen Durchleiter-Bereich hat. Das akustische Signal wird dadurch bei hohen Pegeln gedämpft und bei niedrigeren Pegeln verstärkt. Diese Funktionsweise ist in der Abbildung 5.22 dargestellt, wobei in (a) ein lautes Anregungssi-

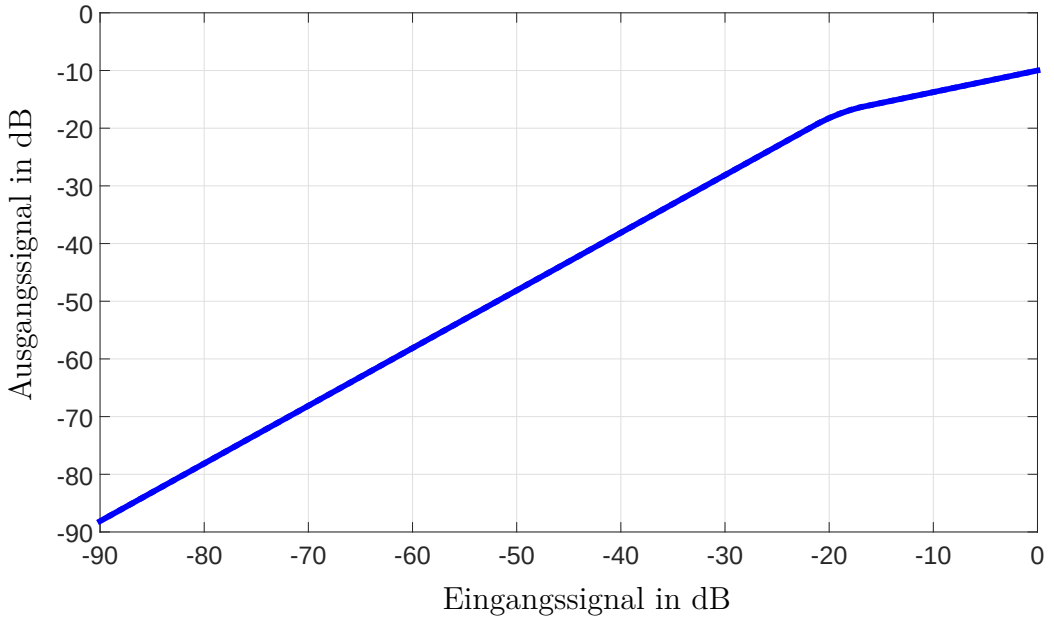
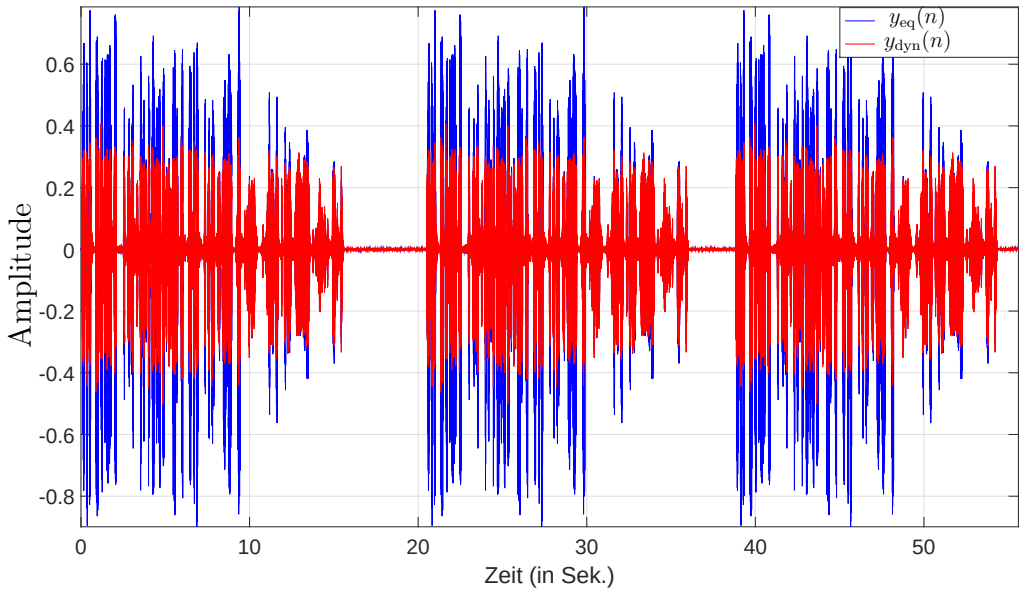
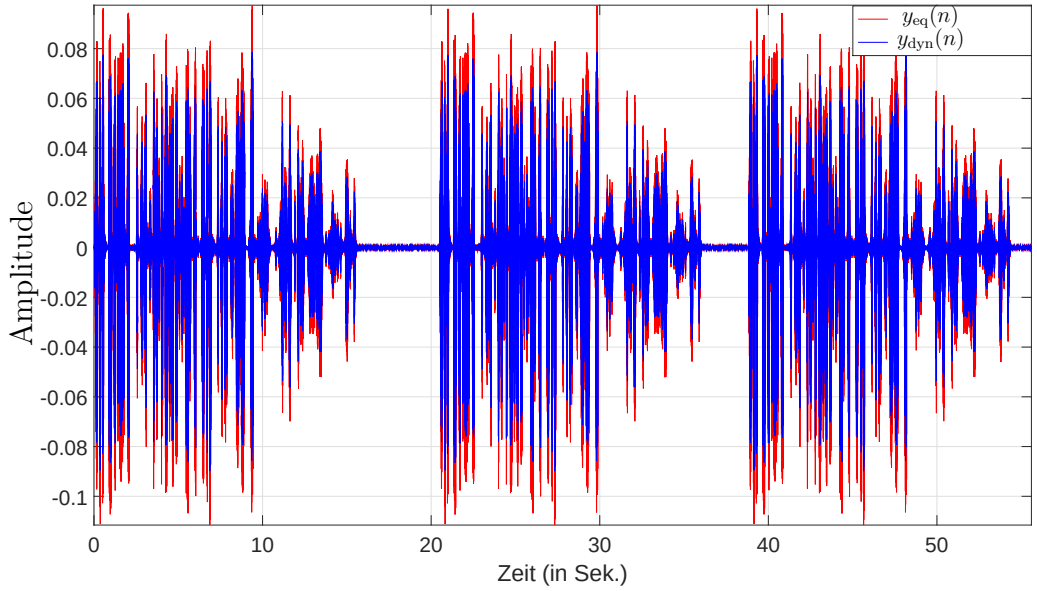


Abbildung 5.21: Dynamik-Kennline der VA-Lautsprecher

gnal und in (b) ein leises Anregungssignal verwendet wurde. Das laute Anregungssignal arbeitet zum Großteil im Kompressor-Bereich und hat auch einige Passagen im Durchleiter-Bereich.



(a)



(b)

Abbildung 5.22: Beispielhafte Dynamikanpassung mit der Dynamik-Kennlinie 5.21 bei (a) lauter und (b) leiser Anregung

Es ist zu sehen, dass die meisten Passagen gedämpft wurden und die leisen Passagen verstärkt wurden. In (b) ist ein leises Anregungssignal gewählt worden, so dass immer der Durchleiter-Bereich wirksam ist, welches ebenfalls sehr gut zusehen ist. Somit können mit

einer Dynamik-Kennlinie verschiedene Effekte erzielt werden. Damit ist das akustische System in der Dynamik anpassbar.

Kapitel 6

Evaluierung

In der Evaluierung wird die Performance von allgemeinen Kommunikationssystemen beschrieben und dafür können verschiedene Maße für die Evaluierung der Systeme genutzt werden [Bro+16]. Aufgrund der verwendeten Algorithmen und der Leistungsbegrenzung können die Kommunikationssysteme der Atemvollschutzmasken in vielen Bereichen mit Freisprechsystemen, Innenraum-Kommunikationssystemen von Fahrzeugen und Hörgeräten verglichen werden. Bei den genannten Applikationen werden bereits Evaluierungen zur Performance-Bestimmung und für Zulassungstests genutzt. Für die Kommunikationssysteme der Atemvollschutzmasken ist ein akustischer Zulassungstest in Nordamerika verpflichtend, welcher eine Speech Transmission Index (STI)-Messung [Ben+08] für die VA-Lautsprecher beinhaltet. Ansonsten sind die Evaluierungen noch nicht in dem Detailgrad wie bei den anderen Applikationen beschrieben und ausgearbeitet. In den nachfolgenden Kapiteln wird ein Überblick gegeben, wie eine Evaluierung für die Kommunikationssysteme der Atemvollschutzmasken aussehen kann. Diese wurde 2016 auf der ITG veröffentlicht [Bro+16]. Zusätzlich wird ein subjektiver Hörtest und die Übertragungsfunktion der Kommunikationseinheit dargestellt.

6.1 Charakteristische Merkmale der Atemvollschutzmaske für die Evaluierung

Die verwendete Sprachaktivitätserkennung [Bro+15] wurde 2015 auf der DAGA veröffentlicht. Die Dämpfung der Atemvollschutzmaske ist der Hauptgrund, warum die akustische Kommunikation schwierig ist. Durch eine aktivierte Kommunikationseinheit kann die Kommunikation deutlich erleichtert werden. Abbildung 6.1 zeigt ein Leistungsdichtespektrum von der Atemvollschutzmaske mit und ohne Kommunikationseinheit, wobei das Mikrofon für die Messung 1 m vor der Maske angeordnet wurde. Bei dieser Maske werden die Frequenzen über 1.5 kHz signifikant gedämpft.
Jede Maske verhält sich hierbei etwas anders, da es hierbei auf den Durchmesser der Sprechmembran, welcher die Resonanzfrequenz bestimmt und auf die Konstruktion der Maske ankommt, wie der Schall aus der Maske austreten kann. Die Verbesserung der Kommunikationseinheit durch die Lautsprecher hat ebenfalls Nachteile, welches sich beispielsweise bei zu großer Verstärkung der VA-Lautsprecher durch eine Rückkopplung äußern kann, bei der die Einheit in einen instabilen Zustand gerät. Ebenso kann die Atemgeräuschreduktion ein zu großes Front-End-Clipping haben, wodurch die ersten Silben der Sprache immer fehlen oder die Detektion fehlerhaft Sprache als Atmen erkennt und somit die Sprache unverständlich wird. Der VA-Lautsprecher kann bei einer zu großen

Verzögerung einen Echo-Effekt bei dem Träger hervorrufen und diesen bei der Kommunikation stören. Um das Verhalten des VA's zu untersuchen wurde ein subjektiver Hörtest durchgeführt. Dafür wird einleitend die Übertragungscharakteristik des VA's der Kommunikationseinheit dargestellt. Nach dem subjektiven Hörtest werden verschiedene Evaluierungsszenarien dargestellt, mit welchen mögliche Fehlverhalten der Algorithmen untersucht werden können.

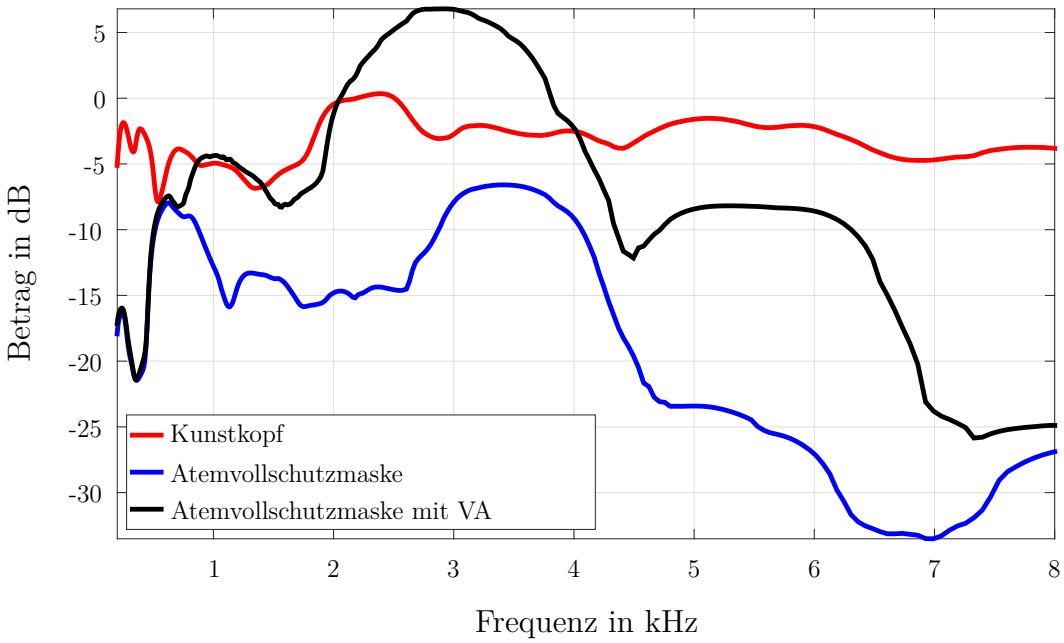
6.2 Subjektiver Hörtest des VA's

6.2.1 Übertragungscharakteristik

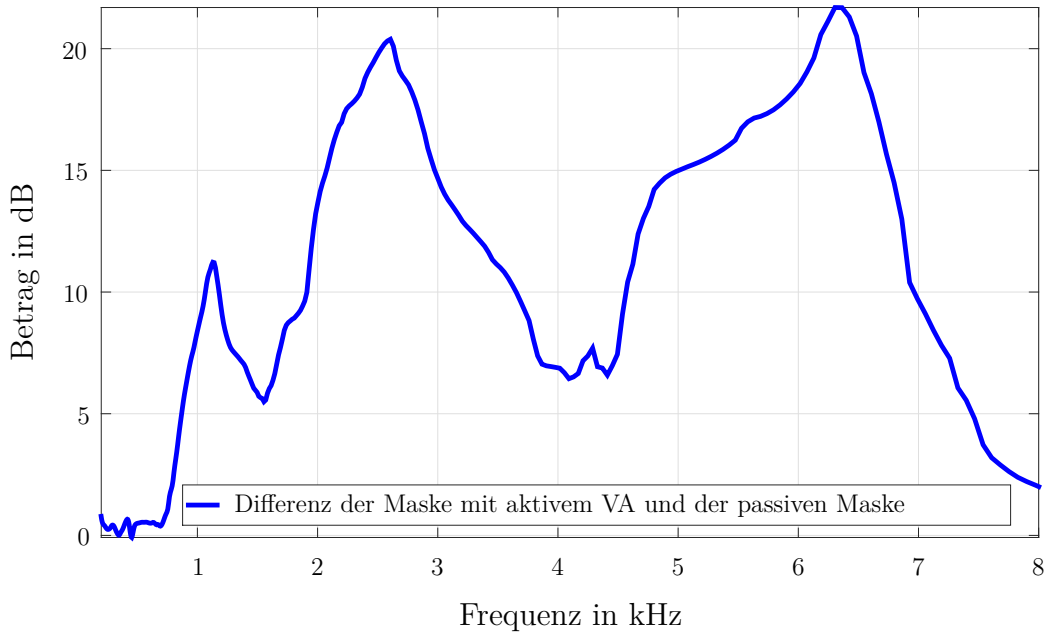
Mit den angewendeten Filtern auf dem VA aus der Abb. 5.16 ergibt sich eine Gesamtübertragungscharakteristik des VA's dargestellt in Abb. 6.1. In dieser Abbildung werden in Teil (a) die Frequenzgänge des Kunstkopfes (rot), der passiven Maske (blau) und der Maske mit aktivem VA der Kommunikationseinheit (schwarz) und in Teil (b) die Differenz zwischen passiver Maske und der Maske mit aktivem VA dargestellt. Bei der Messung befand sich das Mikrofon 1 m vor dem Kunstkopf. Mit einer Maske ohne Kommunikationseinheit kann eine signifikante Dämpfung (blaue Kurve) gemessen werden. Diese beginnt ab ca. 1,5 kHz und die Dämpfung der Frequenzen oberhalb 2 kHz ist sehr stark. Die schwarze Kurve zeigt die Maske mit Kommunikationssystem einschließlich aller vorgestellten Algorithmen, wobei der VA der Kommunikationseinheit oberhalb von 1 kHz eine Verstärkung ausprägt. Diese weist die höchste Verstärkung von ca. 22 dB bei 3 kHz auf. Somit ist eine Verbesserung der Sprachverständlichkeit gegeben. Diese Evaluierung der Steigerung der Sprachverständlichkeit wurde mittels eines *Modified Rhyme Tests* (MRT) durchgeführt, wobei diese Untersuchung den nachfolgenden Regelverstärker beinhaltet.

6.2.2 Subjektiver Hörtest

Die Daten für den modifizierten Reimtest wurden aufgezeichnet und der Test mit zwei Kunstköpfen aufgebaut. Von diesen Köpfen simuliert einer den Sprecher und der andere den Zuhörer [Kon12; ND92]. Die Köpfe sind dabei von Umgebungslautsprechern umgeben, wie in Abb. 6.2 gezeigt. Auf der linken Seite ist ein künstlicher Kopf von GRAS (KEMAR 45) aufgestellt, der den Hörer simuliert. Die zwei Ohrmikrofone erzeugen binaurale Aufnahmen. Links und rechts sind Lautsprecher, welche Umgebungsgeräusche erzeugen, aufgestellt. Gegenüber des GRAS Kunstkopfes ist ein DRÄGER Quaesterkopf [Qua] mit integriertem Lautsprecher platziert, welcher den Sprecher simuliert. Herkömmliche Köpfe wie der KEMAR (aber auch andere) sind etwas zu klein für typische Maskengrößen. Im Test war eine Maske mit einem Kommunikationssystem auf dem Quaesterkopf montiert, so dass im Test die passive Maske und die Maske mit aktivierter Kommunikationseinheit verglichen werden. Die Umgebungsgeräusche werden durch weißes Rauschen dargestellt, welches sehr ähnlich zu einem C-Strahlrohr, einem Schutzventilator und ähnlichen Geräten ist. Das SNR wurde so eingestellt, dass mit der passiven Maske ein $SNR = 0\,\mathrm{dB}$ an den Ohren des Zuhörers erreicht wurde. Bei einer Maske mit aktivem Kommunikationssystem erhöht sich das SNR entsprechend der Übertragungscharakteristik wie in Abb. 6.1 gezeigt. Der modifizierte Reimtest wurde gemäß [ND92] durchgeführt, wobei in diesem Test acht Proben für jede der sieben Reimklassen verwendet werden. Somit wurden 56



(a) Frequenzgang der Atemvollschutzmaske FPS 7000 ohne Kommunikationseinheit (blau), mit Kommunikationseinheit und aktiviertem VA der FPS-COM 7000 (schwarz) und der Kunstkopf ohne Atemvollschutzmaske (rot)



(b) Differenz des Frequenzgangs der Atemvollschutzmaske FPS 7000 mit Kommunikationseinheit und aktiviertem VA der FPS-COM 7000 und der Atemvollschutzmaske FPS 7000 ohne Kommunikationseinheit

Abbildung 6.1

Proben für jede Variante ausgewertet. Der Hörtest wurde von 15 Personen durchgeführt und es wurde eine Fehlerrate für die passive Maske von 23,27 % erreicht. Mit dem akti-

(a) Evaluierungsaufbau ohne Maske

(b) Evaluierungsaufbau mit Maske und Kommunikationseinheit

Abbildung 6.2

viertem Kommunikationssystem wurde ein Fehlerrate von 17,06 % erreicht. Somit erzielt das aktivierte Kommunikationssystem eine Verbesserung um 5,31 Prozentpunkte. Die Teilnehmer haben angegeben, dass Zischlaute besser verstanden werden. Dies kann zum Beispiel sehr wichtig für die Unterscheidung zwischen Wörtern sein, welche sich nur durch das (Plural) „s“ am Ende des Wortes unterscheiden.

6.3 Evaluierungssystem und -szenarien

Bei der Evaluierung der Kommunikationssysteme der Atemvollschutzmasken liegt der Fokus auf vier verschiedenen Evaluierungsszenarien:

1. Eigenwahrnehmung/-störung während der Kommunikation mit den VA-Lautsprechern, wobei der Direktschall als Echo wahrgenommen werden kann und dieses den Träger stört,

2. Kommunikation im Nahfeld durch die VA-Lautsprecher zu anderen Feuerwehrmännern im Trupp oder zu Zivilisten,

3. Kommunikation über Funk durch den Team-Funk zu anderen Feuerwehrmännern im Trupp,

4. Kommunikation über Funk durch das taktische Funkgerät beispielsweise zum Einsatzleiter.

Bei dem Evaluierungssystem ist es wichtig, dass es reproduzierbare und zuverlässige Ergebnisse liefert. Deshalb sollte bei den Tests ein so genannter *h*ead and torso simulator (HATS) genutzt werden. Dieser beinhaltet einen kalibrierten Mundlautsprecher und Ohrmikrofone. Bei der Nahfeld-Kommunikation mit den VA-Lautsprechern wird ein zweiter HATS für den Hörer genutzt, um das Szenario so genau wie möglich nachzubilden. Die Abbildung 6.3 zeigt beispielhaftz zwei verschiedene HATS's, welche im Evaluierungsaufbau genutzt werden.

Abbildung 6.3: *Head and torso simulators* aus unseren Evaluierungsexperimenten.

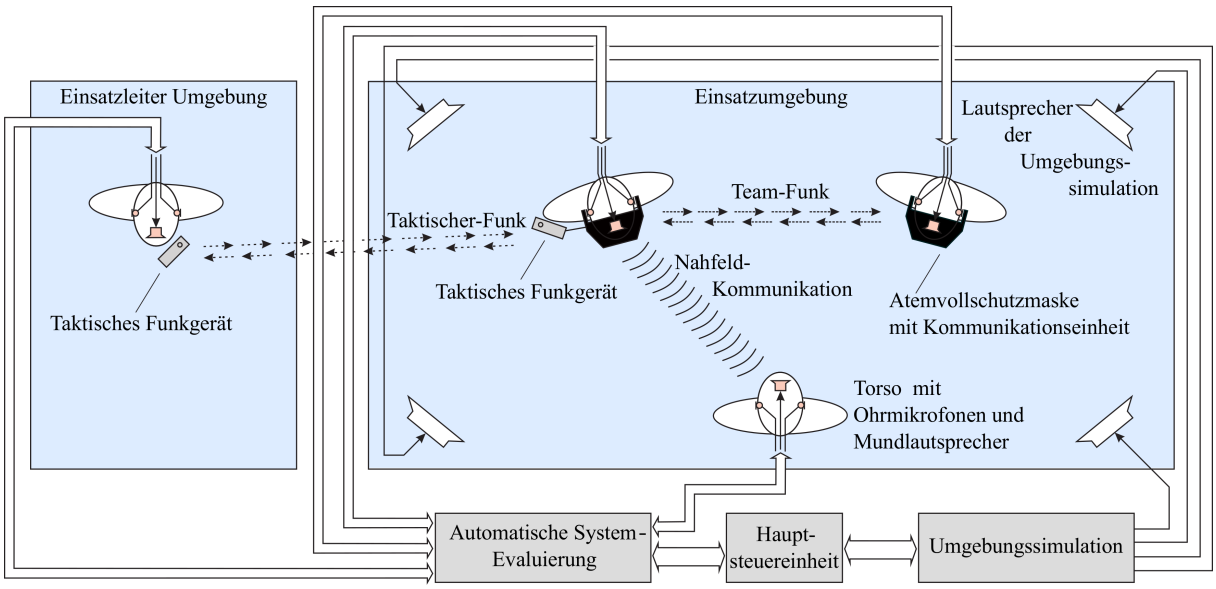


Abbildung 6.4: Überblick der Evaluierungsszenarien für Kommunikationssysteme von Atemvollschutzmasken (verändert nach [Bro+16])

Um die Kommunikationssysteme und dessen Hardware und Algorithmen in möglichst realistischen Szenarien operieren zu lassen, sollte eine Umgebungssimulation genutzt werden. Für diese Umgebungssimulation werden Aufnahmen aus verschiedenen typischen Situationen genutzt, wobei diese Aufnahmen durch zusätzliche Lautsprecher im Evaluierungsszenario wieder gegeben werden. Details über die Umgebungssimulation sind in der Veröffentlichung von Herrn Lüke [Lue+13] zu finden. Zusätzlich kann noch eine Videoleinwand genutzt werden, welche die Personen bei den Hörtests in eine möglichst realistische Umgebung versetzten kann [Lan+15]. Dies hilft dem Zuhörer, sich dem realem Einsatzszenario möglichst nah zu fühlen.

In der Abbildung 6.4 ist ein beispielhafter Aufbau der Evaluierungsszenarien gezeigt. Die Eingangssignale des Evaluierungsmoduls kommen von den Ohrmikrofonen in der Einsatzumgebung und von der Schnittstelle zum taktischen Funkgerät der Kommunikationseinheit. Die Ausgangssignale des Evaluierungsmoduls sind die Mundlautsprecher der einzelnen Köpfe und der taktische Funkgeräte-Eingang der Kommunikationseinheiten. In den nachfolgenden Unterkapiteln sind die einzelnen Evaluierungsszenarien beschrieben.

6.3.1 Eigenwahrnehmung/-störung

Dieses Evaluierungsszenario analysiert die eigene Wahrnehmung des Feuerwehrmannes, wofür der Feuerwehrmann eine Atemvollschutzmaske mit Kommunikationseinheit trägt. Der Feuerwehrmann ist hierbei durch einen Kunstkopf mit Mund und Ohren repräsentiert und der Kunstkopf trägt dabei die Maske. Die Kommunikationseinheit verstärkt die Sprache, welche daraufhin in das Gehör zurückkoppelt. Dieses Signal sollte verglichen mit dem Direktschall nicht mehr als 30 ms verzögert sein, da das Signal der Kommunikationseinheit ansonsten wie ein Echo wahrgenommen werden kann [Wei09]. Gemessen werden kann diese Verzögerung beispielsweise durch eine Kreuzkorrelation. Zusätzlich zu der Ver-

zögerung ist die Sprachqualität und die Sprachverständlichkeit sehr wichtig, so dass die Eigenwahrnehmung nicht beeinträchtigt ist. Dieses kann beispielsweise durch eine spektrale Distanz gemessen werden [The+15], bei welcher das ausgegebene Mundsignal mit dem angekommenen Signal an der Ohren spektral verglichen wird.

6.3.2 Kommunikation mittels der VA-Lautsprecher

Ein weiteres Szenario ist die Übertragung der eigenen Sprache mittels der VA-Lautsprecher zu den Ohren eines Teammitgliedes oder eines Zivilisten. Das Teammitglied kann hierbei noch anders als der Zivilist betrachtet werden, da Erstgenannter einen Helm trägt, welcher das Ohr abschirmen kann und dadurch Dämpfung verursachen kann (siehe Abbildung 6.3). Daher können auch verschiedene Helme unterschiedliche Evaluierungsergebnisse erzielen. Der Zivilist, welcher ein Sanitäter oder eine Opfer sein kann, wird durch eine zusätzliche Person repräsentiert.

Für dieses Szenario wird zusätzlich über die Umgebungslautsprecher Umgebungslärm abgespielt, dies ist in der Abbildung 6.4 dargestellt. Der Zuhörer hat eine definierte Distanz zu dem Sprecher, welche beispielsweise 1,5 m sein kann. Die Umgebungslautsprecher geben typischen Lärm eines Einsatzes wieder, wie das Geräusch eines Feuers oder das eines C-Strahlrohrs.

Bei der Kommunikation mit den VA-Lautsprechern ist für die Verständigung in geräuschbehafteten Umgebungen die Lautstärke und die Sprachverständlichkeit sehr wichtig. Zur Evaluierung der Lautstärke kann das SNR und für die Sprachverständlichkeit eine spektrale Distanz genutzt werden. Zusätzlich kann für die Sprachverständlichkeit der STI [Ben+08, S. 72ff] genutzt werden, welcher bereits Bestandteil des NFPA 1981-2013 Standards für die Kommunikationseinheit von Atemvollschutzmasken ist [Nfpb]. Der STI ist ein objektives Maß, welches die Lautstärke, Hintergrundgeräusche, Reflektionen, Verzögerungen und Verzerrungen im Frequenzbereich mitbetrachtet.

6.3.3 Kommunikation mittels des Team-Funks

Einige Kommunikationssysteme für Atemvollschutzmasken haben ein integriertes Funksystem, welches die Sprache kabellos zu anderen Kommunikationseinheiten sendet, wo sie über die Ohrlautsprecher wiedergegeben wird. Durch solche Systeme wird die Kommunikation im Team deutlich erleichtert und die Kommunikation ist beispielsweise in stark geräuschbehafteten Umgebungen kein Problem. Ein Beispiel ist ein mit Löschschaum gefüllter kleiner Tunnel. Durch den Löschschaum gibt es keine Sichtmöglichkeit und der Direktschall wird stark gedämpft. Hier ist zwar kein Hintergrundgeräusch präsent, aber in diesen Umgebungsbedingungen kann nur mit einem Funksystem zwischen den Teammitgliedern kommuniziert werden. Das Funksystem der Kommunikationssysteme hat im Freifeld eine Reichweite von bis zu 150 m und in Gebäuden ca. 40 m. Somit müssen die Teammitglieder nicht im direktem Sichtfeld für die Kommunikation sein, welches zusätzliche Sicherheit bringt.

Diese Verbesserung und Erleichterung der Kommunikation kann aber nur erreicht werden, wenn die übermittelte Sprache nicht verzerrt ist. Für das Evaluierungsszenario des Team-Funks ist ein Sprecher, mehrere Umgebungslautsprecher und ein Zuhörer notwendig, welches in Abbildung 6.4 zu sehen ist. Das an den Ohren des Zuhörers aufgenommene

Signal wird in Bezug auf das Ausgangssignal des Sprechers analysiert. Dabei wird die Lautstärke und das SNR analysiert, da das Signal laut genug sein muss, allerdings nicht die Ohren beschädigen darf. Eine minimale Lautstärke muss ebenso vorhanden sein, damit die Verständlichkeit in geräuschbehafteter Umgebung gewährleistet wird. Das Umgebungsgeräusch ist wiederum durch die Umgebungslautsprecher realisiert. Die Sprachverständlichkeit kann durch eine Frequenzanalyse bestimmt werden, was beispielsweise durch die spektrale Distanz und/oder das Perceptual Objective Listening Quality Assessment (z.B. POLQA) [Gal15] gegeben ist.

POLQA ist ein standardisiertes Qualitätsmaß, welches in der ITU-T *recommendation* P.863 [P.814] festgehalten ist, wobei hier Tests für Telekommunikationssysteme beschrieben sind. POLQA analysiert das empfangene Signal in Bezug zum übermittelten Signal durch spektrale Variationen, das Geräusch in Sprachpausen, die Verzögerung, die Wiederholung von Rahmen, Lautstärkevariationen usw.. Daher gibt dieses Maß eine gute Aussage über die Sprachqualität des übertragenen Signals. Bezogen auf das zu analysierende System ist die Verzögerung ein sehr wichtiger Punkt, da die Systeme in räumlicher Nähe genutzt werden und sich somit das Funksignal mit dem Direktschall überlagern kann und dies zu einem Echoeffekt führen kann, deswegen muss die Verzögerung des Funksignals so gering wie möglich sein.

6.3.4 Kommunikation mittels des taktischen Funks

Ein dritter Kommunikationspfad kann durch ein taktisches Funkgerät geschaffen werden. Der Feuerwehrmann muss in Gebäuden die Position, den Raum, die Situation usw. über das taktische Funkgerät an den Einsatzleiter berichten, wobei dieser Bericht beim Betreten jedes Raumes erfolgen muss. Dabei steht der Einsatzleiter typischerweise außerhalb des Gebäudes.

Das taktische Funkgerät ist an die Kommunikationseinheit durch ein Kabel oder ein Bluetooth-Modul angebunden und die Sprache wird über das taktische Funkgerät zu einem anderen taktischen Funkgerät übertragen. In der Evaluierung kann das Signal zum und vom taktischen Funkgerät an der Schnittstelle zur Kommunikationseinheit analysiert werden. Falls das taktische Funkgerät mit evaluiert werden soll, kann ein zusätzlicher Kunstkopf in einem separaten Raum positioniert werden (siehe Abbildung 6.4). Für die Evalierung wird über den Kunstkopf ein Testsignal wiedergegeben und das Mikrofonsignal der Kommunikationseinheit wird durch Signalverarbeitung verändert und daraufhin zu dem Interface des taktischen Funkgerätes geleitet. Die andere Kommunikationsrichtung ist vom taktischen Funkgerät zur Kommunikationseinheit.

Für die folgende Betrachtung wählen wir immer die Sichtweise des *Device Under Test* (DUT), welches in diesem Fall die Kommunikationseinheit ist. Das Ausgangssignal ist das Signal von der Kommunikationseinheit zum taktischen Funkgerät und das Eingangssignal ist das empfangene Signal vom taktischen Funkgerät. Im Vergleich zu konventionellen Systemen ist hierbei das Ausgangssignal durch die Charakteristik der Sprechmembran der Maske verändert.

In dem taktischen Funkgerät werden Sprach-Codecs verwendet, welche bei sehr geringen Bitraten operieren können. Daher ist es sehr wichtig, dass das Signal zum taktischen Funkgerät so verständlich wie möglich übertragen wird, da die Codecs die Daten komprimieren und verschiedene Varianten für reine Sprach- und Geräuschpassagen nutzen. Dabei

kann es sein, dass die als Geräusch klassifizierten Passagen gar nicht übertragen werden. Wenn die Sprache durch die Maskencharakteristik verändert wird, kann es bei den Codecs zu Fehlerkennungen der Sprachpassage kommen, welches zu einer sehr schwierigen Verständigung führt. Daher ist eine Evaluierung für das Interface des taktischen Funkgeräts sehr wichtig, so dass die Sprache der Erwartung des Codecs entspricht.

Für das Eingangssignal wird für die Evaluierung eine spektrale Distanz und eine SNR-Analyse angewendet, wobei das Signal von den Ohrhörern abgegriffen wird. Die SNR-Messung wird aufgrund der gleichen Kriterien wie in Kapitel 6.3.3 verwendet. Zusätzlich kann zu den erwähnten Maßen POLQA für die Evaluierung verwendet werden, so dass bei dieser Messung der Pfad mit den Funkgeräten und dessen Funkübertragung gemessen wird.

Das Ausgangssignal wird mit dem künstlichen Mundlautsprecher erzeugt, welches durch das Mikrofon der Kommunikationseinheit aufgenommen wird und im Prozessor verarbeitet wird. Daraufhin wird dieses Signal zum Interface des taktischen Funkgeräts übertragen. Dieses Signal wird dabei durch eine spektrale Distanz und zusätzlich POLQA analysiert, da POLQA noch mehrere Faktoren mit betrachtet.

Die POLQA-Messung ist in dem NFPA-Standard 1802 ab dem Jahr 2019 in Entwurfsversionen und ab dem Jahr 2021 voll im Standard vorhanden [Nfpa]. Damit ist die POLQA-Messung für die Zulassung der Funkgeräte und dessen Zubehör notwendig. Damit wird dargestellt, dass immer mehr Evaluierungsmaße für die akustische Untersuchung in die Standards aufgenommen werden.

6.3.5 Bewertung der Evaluierung

Dieses Kapitel beschreibt die individuelle Bewertungsskala der verschiedenen Messungen, welche in dem Kapitel 6.3 aufgelistet sind. Diese Messungen sind das SNR, die spektrale Distanz (SD), die Verzögerung, POLQA und STI.

POLQA und STI haben feste Bewertungsskalen, welche durch Hörversuche für Standards bestimmt wurden und somit ein Mapping vom objektiven Wert zum subjektiven Empfinden haben. Die Bewertungsskala von POLQA ist gleich der mean opinion scores (MOS) Skala, bei der 1 schlecht und 5 exzellent beschreibt. Eine *Team-Funk*-Kommunikation hat beispielsweise eine POLQA-Bewertung von $3,4$. Dieses Ergebnis ist mit einer elektrischen Anregung und einem elektrischem Abgriff der Signale gemessen und ohne die Betrachtung der Atemvollschutzmaske. Wenn das Setup durch den Kunstkopf mit Maske erweitert wird, sollte die Bewertung schlechter werden. Beispielsweise erzielt das Signal am Ausgang des taktischen Funkgeräts eine Bewertung von $2,9$, wobei die Maske hierbei auf einem Kunstkopf saß. Dieses zeigt den Einfluss der Charakteristik der Maske und der Kommunikationseinheit auf die Bewertung, trotz dass in der Signalverarbeitung bei dieser Messung nur wenige feste Filter angewendet wurden.

Die Bewertungsskala vom STI geht von 0 bis 1 mit folgenden Kategorien: von 0 bis 0,3 sehr schlecht, von 0,3 bis 0,45 schlecht, 0,45 bis 0,6 mittelmäßig, 0,6 bis 0,75 gut und 0,75 bis 1 exzellent. Aktuelle Kommunikationseinheiten erreichen STI-Bewertungen mit dem Pfad der VA-Lautsprecher zu einem Zuhörer zwischen 0,6 und 0,75, welches nach dem NFPA 1981-2013 Standard gemessen wurde.

Für die Verzögerungsmessung wird die Wahrnehmungsschwelle bei Weinzierl [Wei09] beschrieben, wobei der Unterschied zwischen der Wiedergabe und dem Direktschall nicht

größer als 30 ms betragen darf. Bei größeren Unterschieden wird die Wiedergabe als ein sehr störendes Echo wahrgenommen. Diese Schwelle ist natürlich von dem Umgebungslärm, dem Leistungsverhältnis der beiden Quellen und noch mehr Details abhängig. Trotzdem ist die Verzögerungsmessung für das Maß der Eigenstörung mit den VA-Lautsprechern interessant. Die Grenze für das jeweilige Szenario muss mit subjektiven Hörtests überprüft werden und die Bewertung muss für das jeweilige Szenario angepasst werden (Details für Hörtests [P.896]).

Abbildung 6.5 zeigt ein Beispiel eines Qualitätsdiagramms einer Kommunikationseinheit (basierend auf [Ket+08, Kapitel 10.6.1]). Diese Kommunikationseinheit ist mit einer mittleren Qualität klassifiziert. Die vier Hauptevaluierungskategorien sind in Tortenstücken mit durchgezogenen Linien abgebildet. Der Winkel gibt die Gewichtung der individuellen Kategorie an: Somit hat das *Taktische Funkgerät*, der *Team-Funk* und die *VA-Lautsprecher* eine Gewichtung von 30 % und die *Eigenwahrnehmung/-störung* einen geringeren Stellenwert von 10 %, da die Verständigung mit anderen Personen für den Feuerwehrmann eine höhere Priorität hat. Die individuellen Messungen der Kategorien sind mit gestrichelten Linien gekennzeichnet, wobei diese auch durch den Winkel gewichtet sind. Bei der Bewertung ist im Mittelpunkt die schlechteste Bewertung und am Außenrand die beste Bewertung. Somit hat eine Kommunikationseinheit mit einem komplett grün gefüllten Diagramm eine exzellente Qualität. Beim Qualitätsdiagramm müssen noch

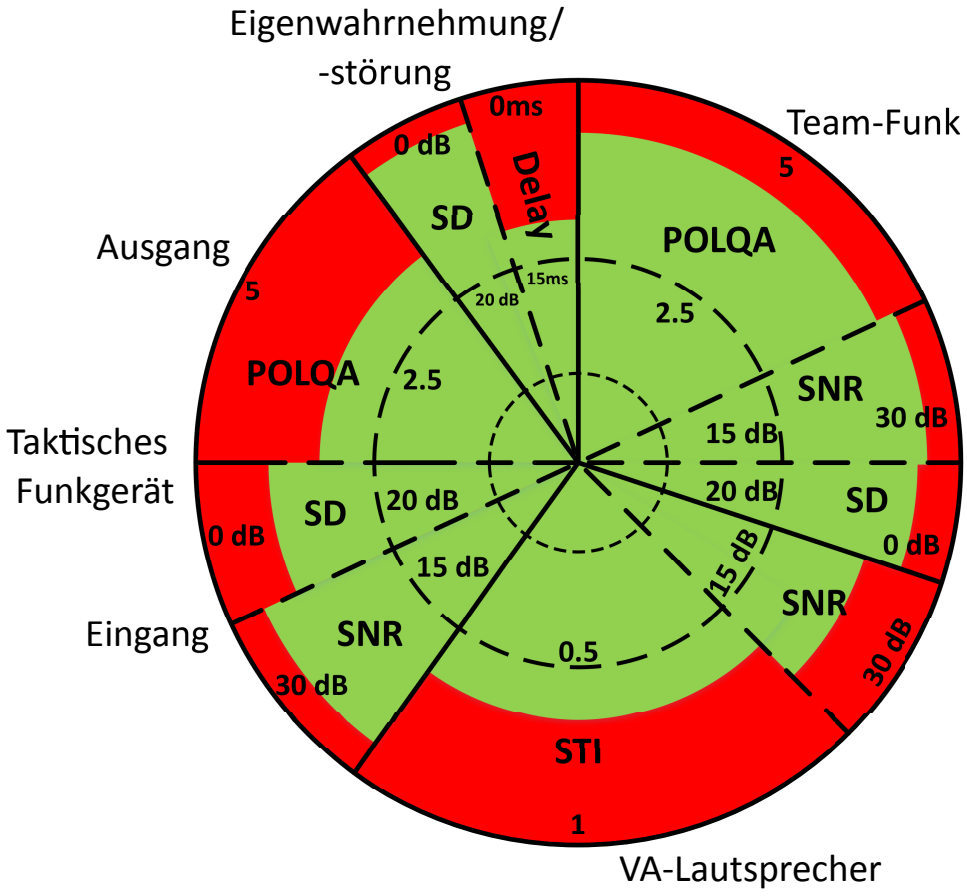


Abbildung 6.5: Beispielhaftes Qualitätsdiagramm einer Kommunikationseinheit von Atemvollschutzmasken (verändert nach [Bro+16]).

die Bewertungsskalen des SNR, der spektralen Distanz und der Verzögerung der einzelnen Szenarien bestimmt werden. Diese Bestimmung kann mittels Hörtests geschehen. Bei STI und POLQA sind die Bewertungen vorgegeben, dabei muss nur das Evaluierungsszenario mit den Maßen vom DUT zu den Messpunkten definiert werden. Eine solche Evaluierung gibt einen sehr guten Überblick über die Qualität der verschiedenen Kommunikationseinheiten in einem Blackbox-Verfahren.

Kapitel 7

Zusammenfassung und Ausblick

Im Rahmen dieser Arbeit wurden Algorithmen zur Verbesserung der Sprachverständlichkeit für Kommunikationssysteme von Atemschutzvollmasken in einer Betrachtung des Gesamtsystems auf einem 16-Bit-Festkomma-DSP umgesetzt. Dies ermöglicht die Nutzung der Algorithmen im Atemschutzeinsatz in der FPS-COM7000 von Dräger. In verschiedenen Hörversuchen konnte dargestellt werden, dass eine Steigerung der Sprachverständlichkeit durch die implementierten Algorithmen möglich war, was durch Einsatzkräfte nach der realen Benutzung auch bestätigt wurde. Der folgende Abschnitt fasst die Algorithmen und die Ergebnisse zusammen, darauf folgt ein Ausblick auf zukünftige Aktivitäten.

7.1 Zusammenfassung

Die vorgestellte Signalverarbeitung des Kommunikationssystems ist darauf ausgelegt, alle Rahmenbedingen aus Kapitel 2 zu betrachten. Bei mehreren Anforderungen für den gleichen Bereich wurde die beste Lösung erarbeitet. In Kapitel 3 wurde eine Übersicht der gesamten Signalverarbeitung gegeben, welche die Mikrofonsignalverbesserung, die Funkgerätesignalverbesserung, die Mischung und Verstärkung und die Nachverarbeitung beinhaltet. Die wichtigsten Signalverarbeitungsblöcke wurden nochmals in eigenständigen Kapitel beschrieben. Die Mikrofonsignalverbesserung wurde in Kapitel 4 und die Nachverarbeitung in Kapitel 5 vorgestellt. Die Mikrofonsignalverbesserung beinhaltet eine Analyse- und Synthesefilterbank, eine Sprachaktivitätsdetektion, eine Rückkopplungskompensation und eine Signaldekorrelation. Die Sprachaktivitätsdetektion erkennt die störenden Atemgeräusche und klassifiziert diese mittels eines Mustererkenners, wobei eine Merkmalsextraktion vor dem Mustererkenner durchgeführt wurde. Dies zeigte sich durch die Untersuchungen in den Testergebnissen als geeignetes Verfahren zur Unterscheidung zwischen Sprache, Geräuschen und Atmen. Die Rückkopplungskompensation ist notwendig, damit eine signifikante Verstärkung bei den vorderen Lautsprechern möglich ist. Durch die starke Korrelation vom Mikrofonsignal zum Lautsprechersignal ist eine Signaldekorrelation des Lautsprechersignals notwendig, um eine möglichst hohe Verstärkung zu erzielen. Die Signaldekorrelation wird mit einem Frequenzversatz umgesetzt. Mit der Dekorrelation und dem Rückkopplungskompensator kann die Verstärkung um mehr als 15 dB erhöht werden. Nach der Mikrofonsignalverbesserung werden die Eingangssignale auf die Ausgangssignale gemischt, bei jedem Ausgangssignal kann die gesamte Nachverarbeitung oder nur Teile dieser angewendet werden. Die Nachverarbeitung beinhaltet einen Exciter, einen Equalizer, einen Regelverstärker und einen Hard-Limiter. Der Exciter erzeugt mittels nichtlinearer Kennlinien Signalanteile, welche durch die Dämpfung der Maske verloren gehen. Der Hörversuch zeigt, dass mit dem Exciter eine höhere Sprachverständlichkeit erzeugt

werden kann. Der Equalizer dient der Entzerrung der Signale, so dass beispielsweise die Einbausituation der Lautsprecher ausgeglichen werden kann. Der Regelverstärker dient zur Dynamikanpassung der Signale, so dass eine möglichst gute Verständlichkeit erzielt wird. Ein Hörversuch mittels eines MRT für das Gesamtsystem zeigt, dass eine Steigerung der Sprachverständlichkeit bei den vorderen Lautsprechern erzielt werden kann. Abschließend wurde ein Evaluierungsszenario vorgestellt, wie eine Bewertung des Gesamtsystems in der Zukunft aussehen könnte, so dass eine Vergleichbarkeit verschiedener Maskenkommunikationssyteme möglich wäre.

7.2 Ausblick

Der Ausblick beinhaltet Verbesserungsmaßnahmen, welche in zukünftigen Arbeiten behandelt werden können. Eine der Verbesserungsmaßnahmen ist eine Rückkopplungsunterdrückung, welche die Rückkopplungen schätzt und diese unterdrücken kann [Wit17]. Mit einiger Anpassung könnte damit eine Unterdrückung der restlichen Rückkopplungen nach der Rückkopplungskompensation erarbeitet werden. Nach dieser Unterdrückung kann somit eine höhere stabile Verstärkung ermöglicht werden. Mit der in dieser Arbeit vorgestellten Rückkopplungskompensation ist eine Vollaussteuerung in Bezug auf das vorhandene elektrische System bereits möglich. Bisher war jedoch die Notwendigkeit nicht gegeben, da die Hardware keine höhere Verstärkung zulässt.

Die vorgestellten Algorithmen dieser Arbeit behandeln nicht die geräuschabhängige Verstärkungkontrolle (NDGC), da das Mikrofon der Kommunikationseinheit zur Schätzung der Umgebungsgeräusche an den VA's und den Ohrlautsprechern ungeeignet ist. Wenn das Kommunikationssystem Außenmikrofone hätte, könnte eine Schätzung des Umgebungsgeräusches an den VA's und den Ohrlautsprechern erfolgen und somit dessen Verstärkung auf Basis der Hintergrundlautstärke erfolgen [Wit17].

Eine weitere Verbesserung wäre die Umarbeitung der passiven Maske, so dass dessen Schallkanal mit der Sprechmembran besser angeordnet wäre und somit der Direktschall der Maske deutlich besser zur Geltung kommt. Damit würde nicht nur die Sprachverständlichkeit bei Masken mit Kommunikationssystemen gesteigert werden, sondern auch beim Nutzer von passiven Masken. Bei den aktiven Kommunikationssystemen würde eine Verlegung des Mikrofons in den Innenraum der Maske einen erheblichen Vorteil bringen. Mit dieser Verlegung wird die Sprache nicht mehr mit dem Resonanzverhalten der Sprechmembran aufgezeichnet. Damit würde ein natürliches Sprachsignal vorliegen. Die Rückkopplungen und Störgeräusche würden durch die Verlegung auch erheblich gedämpft werden, da der Schall von außen durch die Maske gedämpft wird. Somit wären deutlich höhere Verstärkungen möglich und weniger Störgeräusche auf dem Mikrofonsignal vorhanden.

Literaturverzeichnis

Publikationen mit Eigenbeteiligung

[Bro+15] M. Brodersen, A. Volmer, M. Romba und G. Schmidt. *Sprachaktivitätserkennung mittels eines Mustererkenners für Atemschutzmasken*. DAGA, 2015 (siehe S. 3, 22, 34, 87).

[Bro+16] M. Brodersen, T. M. Juengling und G. Schmidt. *Evaluation of Communication Systems for Full-Face Firefighter Masks*. ITG, 2016, S. 1–5 (siehe S. 3, 87, 92, 96).

[Bro+19] Michael Brodersen, Achim Volmer und Gerhard Schmidt. *Signal enhancement for communication systems used by fire fighters*. Bd. 2019. 1. EURASIP Journal on Audio, Speech, und Music Processing, 2019, S. 21 (siehe S. 3, 15, 79).

[Grö+17] B. Gröger, M. Brodersen und G. Schmidt. *Verbesserung der Sprachverständlichkeit für Atemschutzmasken mittels Signalbearbeitung mit nichtlinearen Kennlinien*. DAGA, 2017 (siehe S. 3, 66, 70, 74).

[Mar+17] Carina Marquard, Christin Baasch, Michael Brodersen[1], Oliver Niebuhr und Gerhard Schmidt. *Speak, Think, Act: A phonetic analysis of the combinatorial effects of respiratory mask, physical and cognitive stress on phonation and articulation*. DAGA, Apr. 2017 (siehe S. 3, 17).

Weitere Literatur

[Abl03] S. Ablameyko. *Neural Networks for Instrumentation, Measurement and Related Industrial Applications*. NATO Science Series. IOS Press, 2003 (siehe S. 37).

[Amp15] TPA2006D1 1.45-W MONO Class-D Audio Power Amplifier. *Datasheet*. Texas Instruments, 2015 (siehe S. 13).

[Baa12] C. Baasch. *Verbesserung und Implementierung einer Geräuschschätzung in einem Echtzeitsystem für Anwendungen im Automobilbereich*. Bachelorarbeit, Christian-Albrechts-Universität, Kiel, 2012 (siehe S. 60, 61).

[Ben+08] J. Benesty und Y. Huang. *Springer Handbook of Speech Processing*. Springer handbooks. Springer, 2008 (siehe S. 87, 93).

[Ben+09] J. Benesty, J. Chen, Y. Huang und I. Cohen. *Noise Reduction in Speech Processing*. Springer Topics in Signal Processing. Springer Berlin Heidelberg, 2009 (siehe S. 61).

[Ben+13] J. Benesty, T. Gänsler, D.R. Morgan, M.M. Sondhi und S.L. Gay. *Advances in Network and Acoustic Echo Cancellation*. Digital Signal Processing. Springer Berlin Heidelberg, 2013 (siehe S. 46).

[Ben15] A. Benesch. *Ver- und Entschlüsselungsmethoden im Ersten Weltkrieg. Kryptografieeinsatz an der Westfront und in der Nordsee*. GRIN Verlag, 2015 (siehe S. 32).

[Bis06] C. M. Bishop. *Pattern Recognition and Machine Learning*. 2006 (siehe S. 31, 35).

[Bos+13] D. Boss, K.D. Kammeyer, K. Kristian und A. Dekorsy. *Digitale Signalverarbeitung: Filterung und Spektralanalyse mit MATLAB-Übungen*. Teubner Studienbücher Technik. Vieweg+Teubner Verlag, 2013 (siehe S. 22).

[Bul+16] Philipp Bulling, Klaus Linhard, Arthur Wolf, Gerhard Schmidt, Anne Theiß und Marco Grimm. *Nichtlineare Kennlinien zur Verbesserung der Sprachverständlichkeit in geräuschbehafteter Umgebung*. 2016 (siehe S. 66, 68, 69).

[C55a] *FFT Implementation on theTMS320VC5505, TMS320C5505, and TMS320C5515 DSPs*. Texas Instruments, 2013 (siehe S. 12, 13, 22).

[C55b] *TMS320C55x DSP Library, Programmer's Reference*. Texas Instruments, 2013 (siehe S. 28).

[C55c] *TMS320C55x Optimizing C/C++ Compiler v 4.4*. Texas Instruments, 2013 (siehe S. 12).

[Cha00] Josef Chalupper. „Aural Exciter and Loudness Maximizer: What's Psychoacoustic about -Psychoacoustic Processors?-". In: *Audio Engineering Society Convention 109*. 2000 (siehe S. 66).

[Cle08] T. Cleff. *Deskriptive Statistik und moderne Datenanalyse: eine computergestützte Einführung mit Excel, SPSS und STATA ; [Bachelor geeignet!]* Gabler Lehrbuch. Gabler, 2008 (siehe S. 57).

[Com] *Dräger FPS-COM 7000*. https://www.draeger.com/de_de/Products/FPS-COM-7000 (siehe S. 8).

[Cps] *Dräger CPS 7900*. https://www.draeger.com/de_de/Products/CPS-7900 (siehe S. 9).

[Dav+00] J. Davidson und J. Peters. *Voice over IP - Grundlagen*. Cisco Press. Markt-und-Technik-Verlag, 2000 (siehe S. 24).

[Dav+14] D. Davis und E. Patronis. *Sound System Engineering*. Taylor & Francis, 2014 (siehe S. 55).

[DG+12] Michael Massberg Dimitrios Giannoulis und Joshua D. Reiss. *Digital Dynamic Range Compressor Design - A Tutorial and Analysis*. 2012 (siehe S. 79).

[Dic+08] M. Dickreiter, V. Dittel, W. Hoeg und M. Wöhr. *Handbuch der Tonstudiotechnik*. De Gruyter, 2008 (siehe S. 79).

[Fah+16] L. Fahrmeir, C. Heumann, R. Künstler, I. Pigeot und G. Tutz. *Statistik: Der Weg zur Datenanalyse*. Springer-Lehrbuch. Springer Berlin Heidelberg, 2016 (siehe S. 57, 58).

[Fel+96] J. Feldman und R. Rojas. *Neural Networks: A Systematic Introduction*. Springer Berlin Heidelberg, 1996 (siehe S. 37, 38).

[Fin13] G.A. Fink. *Mustererkennung mit Markov-Modellen: Theorie-Praxis-Anwendungsgebiete*. Vieweg+Teubner Verlag, 2013 (siehe S. 32, 33).

[Fre+16] A. Fred, M. De Marsico und M. Figueiredo. *Pattern Recognition: Applications and Methods: 4th International Conference, ICPRAM 2015, Lisbon, Portugal, January 10-12, 2015, Revised Selected Papers*. Lecture Notes in Computer Science. Springer International Publishing, 2016 (siehe S. 34).

[Fri14] A. Friesecke. *Die Audio-Enzyklopädie: Ein Nachschlagewerk für Tontechniker*. De Gruyter Reference. De Gruyter, 2014 (siehe S. 66).

[Gal15] L.F. Gallardo. *Human and Automatic Speaker Recognition over Telecommunication Channels*. T-Labs Series in Telecommunication Services. Springer Singapore, 2015 (siehe S. 94).

[Ger+97] P. Gerdsen und P. Kröger. *Digitale Signalverarbeitung in der Nachrichtenübertragung*. Digitale Signalverarbeitung in der Nachrichtenübertragung: Elemente, Bausteine, Systeme und ihre Algorithmen. Springer, 1997 (siehe S. 61).

[Gia10] D.C. Giancoli. *Physik: Lehr- und Übungsbuch*. Pearson Studium - Physik. Pearson Studium, 2010 (siehe S. 45).

[Goo+18] I. Goodfellow, Y. Bengio und A. Courville. *Deep Learning. Das umfassende Handbuch: Grundlagen, aktuelle Verfahren und Algorithmen, neue Forschungsansätze*. mitp Verlags GmbH und Company, 2018 (siehe S. 31).

[Gra+15] Simon Graf, Tobias Herbig, Markus Buck und Gerhard Schmidt. *Features for voice activity detection: a comparative analysis*. 2015, S. 91 (siehe S. 22).

[Grü08] D.C. Grünigen. *Digitale Signalverarbeitung: Bausteine, Systeme, Anwendungen: enthält 99 Beispiele und 64 Aufaben; Lösungen, Analyse-, Entwurfs- und Simulationsprogramme*. FO Print und Media, 2008 (siehe S. 77).

[Guo+12] M. Guo, S. H. Jensen, J. Jensen und S. L. Grant. *On the use of a phase modulation method for decorrelation in acoustic feedback cancellation*. 2012, S. 2000–2004 (siehe S. 53).

[Gän+96] Tomas Gänsler, Maria Hansson, Carl-Johan Ivarsson und Göran Salomonsson. *A Double-Talk Detector Based on Ceherence*. 1996 (siehe S. 55).

[Hab99] J. Haber. *Konstruktion und Implementierung eines neuen Verfahrens zur Kompression von Bilddaten*. Utz, Wiss., 1999 (siehe S. 32, 33).

[Hav+08] D.I. Havelock, S. Kuwano und M. Vorlander. *Handbook of Signal Processing in Acoustics*. Springer, 2008 (siehe S. 22).

[Hay96] S.S. Haykin. *Adaptive Filter Theory*. Prentice-Hall information and system sciences series. Prentice Hall, 1996 (siehe S. 46).

[Her+13] E. Hering, R. Martin und M. Stohrer. *Physik für Ingenieure*. Springer-Lehrbuch. Springer Berlin Heidelberg, 2013 (siehe S. 27).

[Hoc07] H. Hochreutener. *Festkomma-Signalprozessor fixed-point DSP*. Zentrum für Signalverarbeitung und Nachrichtentechnik, ZHW, 2007 (siehe S. 77, 78).

[Hof+15] R. Hoffmann und M. Wolff. *Intelligente Signalverarbeitung 2: Signalerkennung*. Springer Berlin Heidelberg, 2015 (siehe S. 31, 35).

[Hol12] M.S. Holambe R.S. und Deshpande. *Advances in Non-Linear Modeling for Speech Processing*. SpringerBriefs in Electrical and Computer Engineering. Springer New York, 2012 (siehe S. 26).

[Hub04] S. Huber. *Untersuchung verschiedener Verfahren zur Grundfrequenzbestimmung mit Einstellung einer Applikation zur Midi-Konvertierung*. Diplom.de, 2004 (siehe S. 7).

[Hut96] H.P. Hutter. *Comparison of Classic and Hybrid HMM Approaches to Speech Recognition Over Telephone Lines*. TIK-Schriftenreihe. vdf, Hochschulverlag AG an der ETH Zürich, 1996 (siehe S. 34).

[Hän+04] E. Hänsler und G. Schmidt. *Acoustic Echo and Noise Control - A Practical Approach*. Wiley, 2004 (siehe S. 21, 46, 47, 49, 50, 61).

[IT01] ITU-T. *Recommendation P.862. Methods for objective and subjective assessment of quality*. 2001 (siehe S. 11, 55).

[IT14] ITU-T. *Recommendation P.863. Methods for objective and subjective assessment of quality*. 2014 (siehe S. 55).

[IT96] ITU-T. *Recommendation P.800. Methods for objective and subjective assessment of quality*. März 1996 (siehe S. 55–57).

[Joh13] J. Johann. *Modulationsverfahren: Grundlagen analoger und digitaler Übertragungssysteme*. Nachrichtentechnik. Springer Berlin Heidelberg, 2013 (siehe S. 53).

[Jun+13] V. Jung und H.J. Warnecke. *Handbuch für die Telekommunikation*. Springer Berlin Heidelberg, 2013 (siehe S. 32).

[Jun13] H.H. Jung. *Neurobasiertes Mass Customizing zur Segmentierung des deutschen PKW-Marktes: Konzeptionelle und methodische Neuausrichtung des Automobilmarketing*. Deutscher Universitätsverlag, 2013 (siehe S. 39).

[Kam+09] K.D. Kammeyer und K. Kroschel. *Digitale Signalverarbeitung: Filterung und Spektralanalyse mit MATLAB-Übungen*. Informations- und Kommunikationstechnik. Vieweg + Teubner, 2009 (siehe S. 22).

[Kea00] S. Keagy. *Integrating Voice and Data Networks*. Cisco Core Series. Cisco Press, 2000 (siehe S. 55).

[Ket+08] F. Kettler und H.-W. Gierlich. „Evaluation of Hands-free Terminals“. In: *Speech and Audio Processing in Adverse Environments*. Hrsg. von E. Hänsler und G. Schmidt. Springer Berlin Heidelberg, 2008 (siehe S. 96).

[Kon12] K. Kondo. *Subjective Quality Measurement of Speech: Its Evaluation, Estimation and Applications*. Signals and Communication Technology. Springer Berlin Heidelberg, 2012 (siehe S. 88).

[Kov+08] B. Kovačević und Ž. Đurović. *Fundamentals of Stochastic Signals, Systems and Estimation Theory with Worked Examples*. Springer, 2008 (siehe S. 61).

[Kuo+01] S.M. Kuo und S.M.K. Bob H. . Lee. *Real-time Digital Signal Processing: Implementations, Applications, and Experiments with the TMS320C55X*. Tsinghua University Press, 2001 (siehe S. 77).

[Lai03] E. Lai. *Practical Digital Signal Processing*. Elsevier Science, 2003 (siehe S. 77).

[Lan+15] R. Landgraf, O. Niebuhr, G. Schmidt, T. John, C. Luecke und A. Theiss. *Von der Straße ins Labor: Die Modifikation der Sprachproduktion bei lauten Fahrgeräuschen*. 2015 (siehe S. 92).

[Lan97] A. Langenmayr. *Sprachpsychologie: Ein Lehrbuch*. Hogrefe Verlag, 1997 (siehe S. 17).

[Ltd01] Aphex Systems Ltd. *Aural Exciter and Optical Big Bottom - Instruction Manual*. 2001 (siehe S. 66).

[Lue+13] C. Lueke, A. Theiss, G. Schmidt, O. Niebuhr und T. John. *Creation of a Lombard Speech Database using an Acoustic Ambiance Simulation with Loudspeakers*. 2013 (siehe S. 92).

[Lük+11] C. Lüke, H. Özer, G. Schmidt, A. Theiß und J. Withopf. *Signal Processing for In-Car Communication Systems*. 5^{th} Biennial Workshop on DSP for In-Vehicle Systems, Kiel, Germany, 2011 (siehe S. 22).

[Mad+00] Andreas Mader, Henning Puder und Gerhard Uwe Schmidt. *Step-size control for acoustic echo cancellation filters – an overview*. Bd. 80. 9. 2000, S. 1697 –1719 (siehe S. 50).

[Mer12] A. Mertins. *Signaltheorie: Grundlagen der Signalbeschreibung, Filterbänke, Wavelets, Zeit-Frequenz-Analyse, Parameter- und Signalschätzung*. SpringerLink : Bücher. Springer Fachmedien Wiesbaden, 2012 (siehe S. 22).

[Mer13] Alfred Mertins. *Signaltheorie. Grundlagen der Signalbeschreibung, Filterbänke, Wavelets, Zeit-Frequenz-Analyse, Parameter- und Signalschätzung*. 3. Aufl. Springer Vieweg, 2013 (siehe S. 55, 60).

[Mey14] M. Meyer. *Signalverarbeitung: Analoge und digitale Signale, Systeme und Filter*. 2014 (siehe S. 22).

[Mil+13] O. Mildenberger und M. Meyer. *Kommunikationstechnik: Konzepte der modernen Nachrichtenübertragung*. Vieweg Praxiswissen. Vieweg+Teubner Verlag, 2013 (siehe S. 53).

[Mir+95] J. Mira und F. Sandoval. *From Natural to Artificial Neural Computation: International Workshop on Artificial Neural Networks, Malaga-Torremolinos, Spain, June 7-9, 1995 : Proceedings*. Lecture Notes in Computer Science. U.S. Government Printing Office, 1995 (siehe S. 37).

[ND92] T. Spillmann N. Dillier. *Deutsche Version der Minimal Auditory Capability (MAC)-Test-Batterie: Anwendungen bei Hörgeräte- und CI-Trägern mit und ohne Störlärm*. 1992 (siehe S. 88).

[Neu12] A. Neubauer. *DFT - Diskrete Fourier-Transformation: Elementare Einführung*. Vieweg+Teubner Verlag, 2012 (siehe S. 21).

[Nfpa] *NFPA 1802-2021, Standard on Two-Way, Portable RF Voice Communications Devices for Use by Emergency Services Personnel in the Hazard Zone*. www.nfpa.org (siehe S. 95).

[Nfpb] *NFPA 1981-2013, Standard on Open-Circuit Self-Contained Breathing Apparatus (SCBA) for Emergency Services*. www.nfpa.org (siehe S. 93).

[P.814] Recommendation ITU-T P.863. *Perceptual objective listening quality assessment*. 2014 (siehe S. 94).

[P.896] Recommendation ITU-T P.800. *Methods for subjective determination of transmission quality*. 1996 (siehe S. 96).

[Pan] *Dräger Panorama Nova*. https://www.draeger.com/de_de/Products/Panorama-Nova (siehe S. 5).

[Pfi+08] B. Pfister und T. Kaufmann. *Sprachverarbeitung*. Springer-Verlag Berlin Heidelberg, 2008 (siehe S. 29).

[Pro13] TMS320C5515 Fixed-Point Digital Signal Processor. *Datasheet*. Texas Instruments, 2013 (siehe S. 12).

[Qua] *Dräger Quaestor 7000*. https://www.draeger.com/de_de/Products/Quaestor-7000 (siehe S. 88).

[Rao+11] K. Rao, D.N. Kim und J.J. Hwang. *Fast Fourier Transform - Algorithms and Applications: Algorithms and Applications*. Signals and communication technology. 2011 (siehe S. 22).

[Rap+07] L.J. Raphael, G.J. Borden und K.S. Harris. *Speech Science Primer: Physiology, Acoustics, and Perception of Speech*. Communication sciences. Lippincott Williams & Wilkins, 2007 (siehe S. 20).

[Rey+11] G.D. Rey und K.F. Wender. *Neuronale Netze: eine Einführung in die Grundlagen, Anwendungen und Datenauswertung*. Aus dem Programm Huber: Psychologie-Lehrbuch. Huber, 2011 (siehe S. 35, 37, 38).

[Sch+12] D. Schönfeld, H. Klimant und R. Piotraschke. *Informations- und Kodierungstheorie*. Vieweg+Teubner Verlag, 2012 (siehe S. 34).

[Sch07] P.M. Schulze. *Beschreibende Statistik*. Oldenbourg, 2007 (siehe S. 30).

[Sch08] M. Schulz. *Voice over IP im Mobilfunk: Untersuchungen von Voice over IP in bestehenden Mobilfunknetzen*. Diplom.de, 2008 (siehe S. 55).

[Sch10] D. Schröder. *Intelligente Verfahren: Identifikation und Regelung nichtlinearer Systeme*. Springer Berlin Heidelberg, 2010 (siehe S. 39).

[Spe10] S. Speidel. *Analyse endoskopischer Bildsequenzen für ein laparoskopisches Assistenzsystem*. KIT Scientific Publ., 2010 (siehe S. 32).

[Ste+37] S. S. Stevens, J. Volkmann und E. B. Newman. *A Scale for the Measurement of the Psychological Magnitude Pitch*. Acoustical Society of America, 1937 (siehe S. 26).

[Str05] T. Strutz. *Bilddatenkompression*. Vieweg Praxiswissen. Vieweg+Teubner Verlag, 2005 (siehe S. 32).

[Tan+08] Z.H. Tan und B. Lindberg. *Automatic Speech Recognition on Mobile Devices and over Communication Networks*. Advances in Computer Vision and Pattern Recognition. Springer London, 2008 (siehe S. 25).

[Tet] *European Telecommunication Standard*. ETS 300 395-1, 1997 (siehe S. 11).

[TFC+16] Low-Power Audio Codec for Portable Audio TLV320AIC34 Four-Channel und Telephony. *Datasheet*. Texas Instruments, 2016 (siehe S. 13, 20).

[The+15] A. Theiß und G. Schmidt. *Spectral Distance Analysis for Quality Estimation of In-Car Communication Systems*. 2015 (siehe S. 93).

[Vol+13] A. Volmer, M. Romba, C. Schmidt und M. Houssem Harbi. *Optimization of Speech Intelligibility for Fire Fighters Full Face Masks*. 2013 (siehe S. 5, 8).

[Wal15] J.F. Walde. *Design Künstlicher Neuronaler Netze: Ein Leitfaden zur effizienten Handhabung mehrschichtiger Perzeptrone*. Wirtschaftswissenschaften. Deutscher Universitätsverlag, 2015 (siehe S. 37).

[Wat+08] R. Watson und O. Downey. *The Little Red Book of Acoustics: A Practical Guide*. Blue Tree Acoustics, 2008 (siehe S. 46).

[Wei09] S. Weinzierl. *Handbuch der Audiotechnik*. VDI-Buch. Springer Berlin Heidelberg, 2009 (siehe S. 27, 58, 92, 95).

[Wei96] W. Weidhaas. *Symmetrische Kabel für zukunftssichere Datennetze: systemunabhängige Basis für die Dienste von heute und morgen ; mit 13 Tabellen*. Kontakt und Studium. Expert-Verlag, 1996 (siehe S. 13).

[Wel+16] T.B. Welch, C.H.G. Wright und M.G. Morrow. *Real-Time Digital Signal Processing from MATLAB to C with the TMS320C6x DSPs, Third Edition*. CRC Press, 2016 (siehe S. 77).

[Wen04] A. Wendemuth. *Grundlagen der stochastischen Sprachverarbeitung*. Oldenbourg, 2004, S. 47–49 (siehe S. 26).

[Wil+16] B.M. Wilamowski und J.D. Irwin. *Intelligent Systems*. ENGnetBASE 2015. CRC Press, 2016 (siehe S. 38).

[Wit+14] J. Withopf, S. Rohde und G. Schmidt. *Application of Frequency Shifting in In-Car Communication Systems*. 2014 (siehe S. 53).

[Wit17] Jochen Withopf. *Signalverarbeitungsverfahren zur Verbesserung der Sprachkommunikation im Fahrzeug*. Shaker Verlag, 2017 (siehe S. 50, 100).

[Zwi82] E. Zwicker. *Psychoakustik (Hochschultext) (German Edition)*. 1. Aufl. Springer, 1982 (siehe S. 27).

[Zöl+02] U. Zölzer, X. Amatriain, D. Arfib, J. Bonada, G. De Poli, P. Dutilleux, G. Evangelista, F. Keiler, A. Loscos, D. Rocchesso u. a. *DAFX - Digital Audio Effects*. John Wiley und Sons, 2002 (siehe S. 77, 79).

[Zöl13] U. Zölzer. *Digitale Audiosignalverarbeitung*. Informationstechnik. Vieweg+ Teubner Verlag, 2013 (siehe S. 79).

Kapitel 8

Anhang

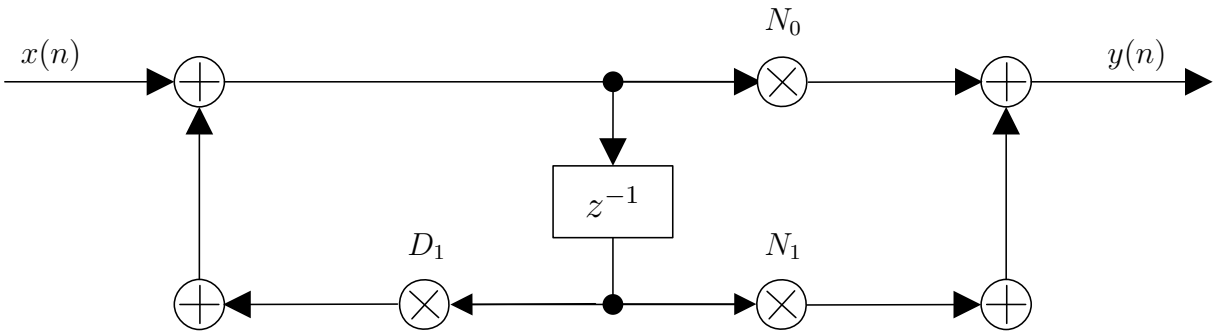


Abbildung 8.1: IIR-Filter erster Ordnung

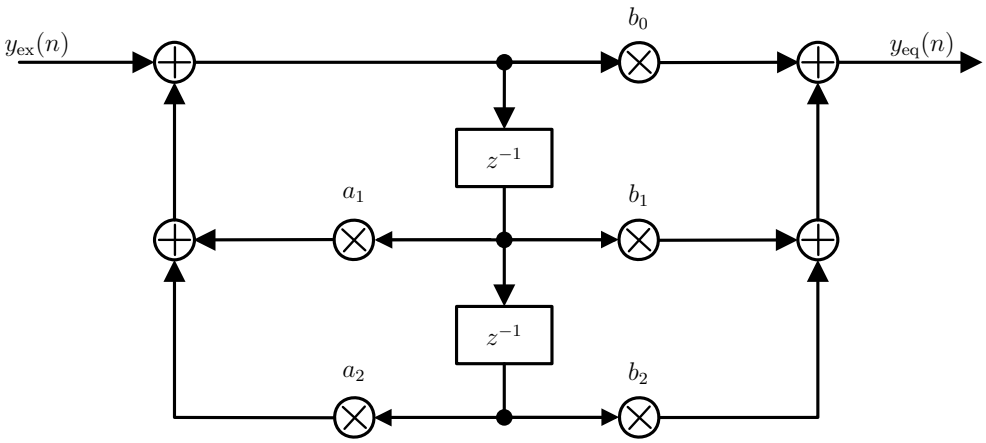


Abbildung 8.2: IIR-Filter 2. Ordnung in der Direkt-Form 2

Erklärung

Hiermit erkläre ich, dass die vorliegende Dissertation nach Inhalt und Form meine eigene Arbeit ist und von mir selbst verfasst worden ist, wobei mir mein Doktorvater Herr Prof. Dr.-Ing. Gerhard Schmidt beratend zur Seite stand. Die Arbeit war weder in Teilen noch im Ganzen Bestandteil eines früheren Prüfungsverfahrens und ist an keiner anderen Stelle zur Prüfung eingereicht. Der Inhalt der Arbeit wurde in Teilen in meinen wissenschaftlichen Publikationen veröffentlicht. Dies ist in der Arbeit entsprechend vermerkt. Die Arbeit ist nach bestem Wissen und Gewissen konform mit den Regeln guter wissenschaftlicher Praxis, welche durch die Deutsche Forschungsgemeinschaft festgelegt sind.

Ort Datum Michael Brodersen